



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

FÁBIO PIRES ITTURRIET

**PRESERVAÇÃO PERCEPTUALMENTE RELEVANTE DA DIFERENÇA DE
TEMPO INTERAURAL EM APARELHOS AUDITIVOS BINAURAIS**

FLORIANÓPOLIS

2019

Fábio Pires Itturriet

**PRESERVAÇÃO PERCEPTUALMENTE RELEVANTE DA DIFERENÇA DE
TEMPO INTERAURAL EM APARELHOS AUDITIVOS BINAURAIIS**

Tese submetida ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina para a obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Prof. Márcio Holsbach Costa, Dr.

Florianópolis

2019

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Itturriet, Fábio Pires

Preservação perceptualmente relevante da diferença de tempo interaural em aparelhos auditivos binaurais / Fábio Pires Itturriet ; orientador, Márcio Holsbach Costa, 2019. 159 p.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Elétrica, Florianópolis, 2019.

Inclui referências.

1. Engenharia Elétrica. 2. Aparelhos auditivos binaurais. 3. Redução de ruído. 4. Diferença de tempo interaural. I. Costa, Márcio Holsbach . II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

Fábio Pires Itturriet

**Preservação perceptualmente relevante da diferença de tempo interaural em aparelhos
auditivos binaurais**

O presente trabalho em nível de doutorado foi avaliado e aprovado por banca examinadora
composta pelos seguintes membros:

Profa. Rosângela Fernandes Coelho, Phd.
Instituto Militar de Engenharia

Prof. Júlio Apolinário Cordioli, Dr.
Universidade Federal de Santa Catarina

Prof. José Carlos Moreira Bermudez, Phd.
Universidade Federal de Santa Catarina

Prof. Bruno Catarino Bispo, Dr.
Universidade Federal de Santa Catarina

Certificamos que esta é a **versão original e final** do trabalho de conclusão que foi julgado
adequado para obtenção do título de doutor em Engenharia Elétrica.

Prof. Bartolomeu Ferreira Uchôa-Filho, Phd.
Coordenador do Programa

Prof. Márcio Holsbach Costa, Dr.
Orientador

Florianópolis, 30 de Agosto de 2019.

Imagination is more important than knowledge. Knowledge is limited whereas imagination embraces the entire world, stimulating progress, giving birth to evolution.

(Albert Einstein, 1929)

AGRADECIMENTOS

Em primeiro lugar agradeço a Deus por ter me dado forças e inspiração para a realização do presente trabalho.

Agradeço ao meu pai José Luiz Itturriet, minha mãe Diva Eunice Itturriet e minha irmã Ana Paula Itturriet pelo constante apoio nesse período.

Agradeço também aos amigos Ciro Eder, Diego Carmo, Wemerson Parreira, Renata Borges, Delvanei Bandejas, Marcelo Balin, George Gavioli, Thiago Araújo, Luciano Gentilini, Luciano Pinto, Emílio Roth e Ellen Souza pelas conversas e churrascos que ajudaram a deixar esse período um pouco mais leve.

Por fim, agradeço ao meu orientador Márcio Holsbach Costa pela excelente orientação e amizade demonstrada nesse período.

RESUMO

A audição é um sentido fundamental para a comunicação entre seres humanos, como também para alerta em situações de risco e para prover consciência espacial. Essa consciência consiste na capacidade de localizar a posição das fontes sonoras que compõem uma determinada cena acústica. A capacidade de seletividade espacial dos seres humanos permite a ênfase de informação sonora de interesse, em detrimento de outras, mesmo quando submetida a cenas acústicas complexas. Esse fenômeno é conhecido na literatura como efeito *cocktail party*. Nesse processo, o cérebro utiliza as diferenças acústicas (pistas), geradas pelo ângulo de incidência e pelo conjunto pina/cabeça, na forma de atrasos e atenuações entre os sons que chegam até as orelhas. O atraso entre os sinais acústicos entre ambas orelhas é conhecido como diferença de tempo interaural e a atenuação como diferença de nível interaural. Quando essas pistas sofrem modificações indesejadas, uma imagem virtual equivocada da fonte acústica é gerada, comprometendo a percepção da cena acústica original. Indivíduos que apresentam dificuldade para escutar, usualmente utilizam aparelhos auditivos. Esses dispositivos, além de compensar as perdas devem oferecer conforto acústico ao usuário (redução de ruído) e preservar a cena acústica original. Quando essas perdas acometem ambas as orelhas, aparelhos auditivos binaurais são indicados, pois compartilham informações entre si, e por isso ampliam o acesso à informação espacial melhorando o desempenho do sistema e a satisfação do usuário. Uma das técnicas de redução de ruído mais utilizadas na literatura para aparelhos auditivos binaurais é o filtro de Wiener Multicanal, que embora eficiente em termos de redução de ruído, modifica as pistas binaurais das fontes sonoras. Por esse motivo, várias soluções foram propostas na literatura visando restaurar a localização original das fontes sonoras. Algumas delas, entretanto, comprometem sobremaneira a redução de ruído, enquanto outras não são efetivas na manutenção da percepção espacial do usuário. A partir do exposto, o presente trabalho apresenta uma nova proposta para preservar a localização original de fontes sonoras pontuais em baixas frequências. Critérios objetivos e experimentos psicoacústicos são apresentados para a demonstração de seu desempenho.

Palavras-chave: Aparelhos auditivos binaurais. Redução de ruído. Diferença de tempo interaural.

ABSTRACT

Hearing is fundamental for communication between human beings, as well as for alert and to provide spatial consciousness. The consciousness consists on the capacity of localizing sound sources in complex acoustic scenes. The spatial selective capacity of human beings permits to emphasize the acoustic information of interest in detriment of other sound sources, even under complex acoustic situations. This phenomenon is known as cocktail party effect. In this process, the brain uses the acoustic differences (cues), generated by the angle of incidence of the arriving sound, as well as by the pinna/head influence, in the form of delays and attenuations between the sounds that arrive at each ear. The resulting delay between the acoustic signals at each ear is called interaural time difference, and the attenuation is the interaural level difference. In case of distortions of these cues, an equivocated virtual image of the acoustic source is generated, compromising the subjective perception of the original acoustic scene. Hearing impaired people are usually benefited by wearing hearing aids. These gadgets aim not only compensation of the hearing losses, but also to increase the acoustic comfort (noise reduction) and to preserve the original acoustic scene. In case of bilateral losses, binaural hearing aids are indicated, since they share information, increasing spatial information, the system performance, as well as the user satisfaction. One of the most applied noise reduction techniques for binaural hearing aids is the Binaural Multichannel Wiener Filter. However, despite being efficient for noise reduction it changes the binaural cues of the acoustic sources. In this way, many works in the literature approach such problem. Some of them, however, compromise the amount of noise reduction or are not effective in restoring the true spatial perception of the user. Due to these facts, this work presents a new method for spatial preservation of low frequency acoustic scenes. A theoretical basis, objective metrics and psychoacoustics experiments are provided to support and demonstrate its performance.

Keywords: Binaural hearing aids. Noise reduction. Interaural time difference.

LISTA DE FIGURAS

Figura 1 – Taxa de adesão ao uso de aparelhos auditivos por idade.	22
Figura 2 – Satisfação dos usuários de aparelhos auditivos em cenas acústicas complexas.	23
Figura 3 – Ocorrência anual de publicações sobre técnicas de redução de ruído e estratégias auxiliares de preservação de localização da fonte de ruído direcional.....	26
Figura 4 – Representação da anatomia do sistema auditivo humano.....	32
Figura 5 – ITD em função do azimute da fonte sonora.	34
Figura 6 – Variação da ILD com o azimute da fonte direcional.	35
Figura 7 – Mapeamento entre azimute e as pistas ITD e ILD.....	35
Figura 8 – Exemplificação do modelo <i>Jeffress</i>	37
Figura 9 – Mapeamento de um azimute para uma IPD.	39
Figura 10 – Modificação da teoria duplex segundo a interpretação ecológica.	41
Figura 11 – Indivíduo localizando fonte sonora em campo aberto (esquerda) e lateralizando fonte sonora com o auxílio de fones de ouvido (direita).	43
Figura 12 – Relação entre os diferentes graus de perdas auditivas com sons encontrados no cotidiano.	45
Figura 13 – Ilustração da cóclea com ênfase nas células ciliadas separadas por bandas de frequência.	47
Figura 14 – Ilustração de um aparelho BTE modelo <i>Dalia</i> da fabricante <i>Phonak</i>	49
Figura 15 – Ilustração de um aparelho ITE.	49
Figura 16 – Diagrama em blocos do problema de filtragem estatística.	55
Figura 17 – Diagrama em blocos de um filtro de Wiener com N coeficientes.....	56
Figura 18 – Utilização de um detector de fala para selecionar a densidade espectral de potência a ser estimada.....	61
Figura 19 – Configuração de um aparelho auditivo binaural.	62
Figura 20 – Problema do MWF na localização da fonte de ruído.....	65
Figura 21 – Metodologia experimental utilizada no trabalho.....	79
Figura 22 – Cenário acústico virtual gerado para os experimentos.....	80
Figura 23 – Exemplo de geração de cenário acústico com uma fonte de fala e uma fonte de ruído.....	81
Figura 24 – Comportamento da métrica Δ ITD do ruído em 50 realizações para a configuração S_0N_{-60}	82

Figura 25 – Comportamento da métrica ΔITD do ruído em 50 realizações para a configuração S_0N_{60} .	82
Figura 26 – Resultados psicoacústicos médios com arquivos de áudio do ruído não processado. Voluntários: 1 (o), 2 (Δ), 3 (∇), 4 (\square), 5 (\diamond). Resultados referentes a -60° em vermelho, 0° em branco e $+60^\circ$ em azul.	84
Figura 27 – Resultados psicoacústicos médios obtidos para o ruído processado pela solução ótima da equação (52) para $\beta = 0$ (J_{MWF}). Voluntários: 1 (o), 2 (Δ), 3 (∇), 4 (\square), 5 (\diamond). Resultados referentes a -60° em vermelho e $+60^\circ$ em azul.	85
Figura 28 – Resultados psicoacústicos médios obtidos para o ruído processado pela solução ótima da equação (52) para $\beta = 10^7$. Voluntários: 1 (o), 2 (Δ), 3 (∇), 4 (\square), 5 (\diamond). Resultados referentes a -60° em vermelho e $+60^\circ$ em azul.	85
Figura 29 – Função de densidade de probabilidade marginal para a fase de $v_{L,ref}/v_{R,ref}$ como função de $ \rho = \mathbb{E} \{ v_{L,ref}/v_{R,ref}^* \} / (\sigma_{v_L} \sigma_{v_R})$.	89
Figura 30 – ΔMSC da técnica apresentada em (52) para a configuração S_0N_{-60} .	92
Figura 31 – ΔMSC da técnica apresentada em (52) para a configuração S_0N_{60} .	92
Figura 32 – Comparação do ΔMSC entre a técnica da equação (52) (vermelho) com a técnica proposta em (93) com $\gamma = 0,05$ (magenta) e $\gamma = 1$ (verde) para a configuração S_0N_{-60} .	95
Figura 33 – Comparação do ΔMSC entre a técnica da equação (52) (azimute) com a técnica proposta em (93) com $\gamma = 0,05$ (magenta) e $\gamma = 1$ (verde) para a configuração S_0N_{60} .	95
Figura 34 – ΔMSC para a função custo J_{IC}^γ , equação (100) para cenário acústico S_0N_{-60} .	100
Figura 35 – ΔMSC para a função custo J_{IC}^γ , equação (100) para cenário acústico S_0N_{60} .	100
Figura 36 – Resumo sobre os resultados dos experimentos.	101
Figura 37 – Cena acústica virtual gerada para os experimentos.	102
Figura 38 – Classificação dos trechos de voz em fala+ruído ($VAD = 1$) ou apenas ruído ($VAD = 0$).	103
Figura 39 – Periodogramas das fontes de fala (vermelho) e ruído (preto).	104
Figura 40 – Comportamento de ΔITD com a variação do parâmetro de controle β da função custo J_{MSC} com dois valores fixos de γ para S_0N_{-60} (a) e S_0N_{60} (b).	106
Figura 41 – Comportamento de ΔITD com a variação do parâmetro de controle β da função custo J_{IC} para S_0N_{-60} (a) e S_0N_{60} (b).	107
Figura 42 – Função custo J_{IC} . Escolha do menor β que minimiza ΔITD e ΔMSC para o cenário acústico S_0N_{-60} .	108

Figura 43 – Função custo J_C . Escolha do menor β que minimiza ΔITD e ΔMSC para o cenário acústico S_0N_{60} .	108
Figura 44 – Resultados de conforto acústico para o cenário acústico S_0N_{-60} .	110
Figura 45 – Resultados de conforto acústico para o cenário S_0N_{60} .	110
Figura 46 – MOS-LQOR referentes às técnicas analisadas para o cenário acústico S_0N_{-60} .	111
Figura 47 – MOS-LQOL referentes às técnicas analisadas para o cenário acústico S_0N_{60} .	111
Figura 48 – Resultados de Inteligibilidade das técnicas para o cenário acústico S_0N_{-60} .	112
Figura 49 – Resultados de Inteligibilidade das técnicas para o cenário acústico S_0N_{60} .	113
Figura 50 – Resultados subjetivos para o cenário acústico S_0N_{-60} .	114
Figura 51 – Resultados subjetivos para o cenário acústico S_0N_{60} .	114
Figura 52 – SNR na orelha esquerda (a) e direita (b) para técnica 1 (azul) e técnica 3 (vermelho) e S_0N_{60} (linha contínua) e S_0N_{30} (linha tracejada) nos cenários (i) J_T para S_0N_{60} ; (ii) J_C para S_0N_{60} ; (iii) J_T para S_0N_{30} ; (iv) J_C para S_0N_{30} .	118
Figura 53 – $\Delta ISNR$ nas orelhas esquerda (a) e direita (b) para MWF-ITD (azul) e MWF-IC (vermelho). Cenários S_0N_{60} (linha contínua) e S_0N_{30} (linha tracejada): (i) J_T para S_0N_{60} ; (ii) J_C para S_0N_{60} ; (iii) J_T para S_0N_{30} ; (iv) J_C para S_0N_{30} .	119
Figura 54 – MOS-LQO nas orelhas esquerda (a) e direita (b) para as técnicas MWF-ITD (azul) e MWF-IC (vermelho) nos cenários S_0N_{60} (linha contínua) e S_0N_{30} (linha tracejada): (i) J_T para S_0N_{60} ; (ii) J_C para S_0N_{60} ; (iii) J_T para S_0N_{30} ; (iv) J_C para S_0N_{30} .	119
Figura 55 – ΔITD do ruído (a) e ΔMSC do ruído para técnica 1 (azul) e técnica 3 (vermelho) e S_0N_{60} (linha contínua) e S_0N_{30} (linha tracejada) nos cenários (i) J_T para S_0N_{60} ; (ii) J_C para S_0N_{60} ; (iii) J_T para S_0N_{30} ; (iv) J_C para S_0N_{30} .	120
Figura 56 – Espectrograma dos sinais na orelha direita para o cenário S_0N_{60} : (a) fala limpa; (b) fala contaminada; (c) fala contaminada processada pelo MWF; (d) fala contaminada processada pelo MWF-ITD; (e) fala contaminada processada pelo MWF-IC.	124
Figura 57 – Classificação dos resultados obtidos na fase de treinamento para os 15 voluntários com o ruído com frequência abaixo de 1,5 kHz.	124
Figura 58 – Boxplot modificado para o experimento psicoacústico 2. Percepção do azimute da fonte de ruído. (a) S_0N_{30} (cima), e (b) S_0N_{60} (baixo): Ruído processado por MWF-IC (azul), MWF-ITD (vermelho), MWF (verde), e ruído não processado (amarelo).	125
Figura 59 – Boxplot modificado para o experimento psicoacústico 2. Percepção do azimute da fonte de ruído. (a) S_0N_{-30} (cima), e (b) S_0N_{-60} (baixo): Ruído processado por MWF-IC (azul), MWF-ITD (vermelho), MWF (verde), e ruído não processado (amarelo).	126

Figura 60 – Inversões de lateralização obtidas nos experimentos psicoacústicos para técnica MWF-ITD: 5 voluntários e 15 tentativas para cada.	128
Figura 61 – Magnitude da coerência interaural: (a) ruído não processado (quadrado amarelo); (b) ruído processado pelo MWF (círculo verde); (c) ruído processado pelo MWF-ITD (triângulo vermelho); (d) ruído processado pelo MWF-IC (asterisco azul).....	130
Figura 62 – Cenário reverberante <i>Office I</i>	131
Figura 63 – Experimento subjetivo com 10 voluntários para o cenário S_0N_{-60} em condições a) anecoicas e b) reverberantes.	132
Figura 64 – Experimento subjetivo com 10 voluntários para o cenário S_0N_{60} em condições a) anecoicas e b) reverberantes.	132
Figura 65 – Diagrama do esquema aplicado nos experimentos psicoacústicos.	145
Figura 66 – Janela da etapa de aprendizado.	145
Figura 67 – Janela da etapa de treinamento.....	146
Figura 68 – Janela da etapa de avaliação.....	147

LISTA DE TABELAS

Tabela 1 – Avaliação da solução que visa preservar a ITD da fonte de ruído.	28
Tabela 2 – Mapeamento de azimute para ITD, banda útil e banda crítica para localização de fontes direcionais baseados no tamanho médio da cabeça humana.	40
Tabela 3 – Mapeamento MOS-LQO na escala ACR.	52
Tabela 4 – Quadro comparativo entre diferentes variações da técnica de MWF.	77
Tabela 5 – Resumo das técnicas avaliadas no experimento 1 em campo livre.	105
Tabela 6 – Resumo das soluções exploradas no experimento 1.	109
Tabela 7 – Mapeamento dos resultados para MOS-LQO para ACR.	112
Tabela 8 – Resumo da análise estatística das técnicas avaliadas com experimentos subjetivos.	116
Tabela 9 – Resumo das soluções exploradas no experimento 2.	117
Tabela 10 – Medidas objetivas para o sinal de fala contaminada.	117
Tabela 11 – Medidas objetivas para S_0N_{30} : $\alpha_T = 3 \times 10^3 (J_T)$, $\alpha_C = 0,8 (J_C)$	121
Tabela 12 – Medidas objetivas para S_0N_{60} : $\alpha_T = 400 (J_T)$, $\alpha_C = 0,4 (J_C)$	122
Tabela 13 – Fator de ponderação para uma perda de redução de ruído de no máximo 15% comparado com a técnica MWF.	123
Tabela 14 – Experimento psicoacústico: Média amostral do azimute (\bar{x}_i), desvio padrão (s_i) e mediana (\tilde{x}_i) tal que $i = \{ Raw, W, T, C \}$, significando respectivamente: não processado ..	127
Tabela 15 – Experimento psicoacústico: Média entre as diferenças absolutas (\bar{x}_i) e mediana (\tilde{x}_i) dos azimutes para MWF-ITD (T) e MWF-IC (C) quando comparados com o azimute percebido do ruído RAW (R).	127

LISTA DE ABREVIATURAS E SIGLAS

ACR *Absolute Category Rating*
ATF *Acoustic Transfer Function*
DOA *Directional of Arrival*
DSP *Digital Signal Processing*
DFT *Discrete Fourier Transform*
HRTF *Head-Related Transfer Function*
ILD *Interaural Level Difference*
IPD *Interaural Phase Difference*
ISTF *Inverse Short-Time Fourier Transform*
ITD *Interaural Time Difference*
ITU *International Telecommunication Union*
LQO *Listening Quality Objective*
MOS *Mean Opinion Score*
MSC *Magnitude Squared Coherence*
PESQ *Perceptual Evaluation of Speech Quality*
RTF *Relative Transfer Function*
SDW *Speech Distortion Weighted*
SPL *Sound Pressure Level*
STFT *Short-Time Fourier Transform*
VAD *Voice Activity Detector*

SUMÁRIO

1	INTRODUÇÃO	18
1.1	MOTIVAÇÃO	20
1.1.1	Tendências no mercado mundial de aparelhos auditivos	20
1.1.2	Fatores de risco para a saúde correlacionados com perdas auditivas	24
1.1.3	Técnicas de redução de ruído com preservação de lateralização.....	24
1.2	PROBLEMA DE PESQUISA	28
1.3	OBJETIVOS	29
1.3.1	Objetivos gerais.....	29
1.3.2	Objetivos específicos.....	29
1.4	ORGANIZAÇÃO DO TEXTO	29
2	FUNDAMENTAÇÃO TEÓRICA.....	31
2.1	ANATOMIA DO SISTEMA AUDITIVO HUMANO	31
2.2	PROCESSAMENTO BINAURAL DE FONTES SONORAS DIRECIONAIS ..	33
2.2.1	Relação entre azimute e as pistas ITD e ILD	33
2.2.2	Teoria Duplex.....	36
2.2.3	Modelo <i>Jeffress</i> para ITD	37
2.2.4	Limite anatômico imposto pelo tamanho da cabeça humana no uso da ITD	38
2.2.5	Peso da ILD na banda dominada pela ITD.....	41
2.2.6	Localização versus lateralização de fontes sonoras direcionais	42
2.3	GRAUS E TIPOS DE PERDAS AUDITIVAS.....	43
2.4	PRINCIPAIS MODELOS DE APARELHOS AUDITIVOS.....	47
2.4.1	Aparelhos BTE (<i>Behind-The-Ear</i>).....	48
2.4.2	Aparelhos ITE (<i>In-The-Ear</i>)	49
2.4.3	Aparelhos Binaurais	49
2.5	MÉTRICAS DE AVALIAÇÃO DA FALA PROCESSADA.....	50

2.5.1	Avaliação objetiva de qualidade.....	51
2.5.2	Avaliação objetiva de inteligibilidade	52
2.6	MÉTRICAS DE AVALIAÇÃO DA PRESERVAÇÃO DA LOCALIZAÇÃO ...	52
2.6.1	Erro de ITD	53
2.6.2	Erro de ILD	53
2.6.3	Avaliação psicoacústica	53
3	FILTRAGEM WIENER MULTICANAL PARA REDUÇÃO DE RUÍDO..	55
3.1	FILTRO DE WIENER MONOCANAL PARA REDUÇÃO DE RUÍDO	59
3.2	FILTRO DE WIENER MULTICANAL PARA REDUÇÃO DE RUÍDO	61
3.3	EFEITOS DA FILTRAGEM WIENER MULTICANAL NAS PISTAS BINAURAS DA FALA E RUÍDO	65
3.4	FILTRAGEM WIENER MULTICANAL COM TÉCNICAS DE PRESERVAÇÃO DAS PISTAS BINAURAS DA FONTE DE RUÍDO	66
3.4.1	MWF-SDW (<i>Speech Distortion Weighted</i>).....	66
3.4.2	MWF- η (<i>Partial Estimation of Noise Component</i>).....	67
3.4.3	MWF-ILD (<i>Interaural Level Difference</i>)	67
3.4.4	MWF-ITD (<i>Interaural Time Difference</i>).....	70
3.4.5	MWF-ITF (<i>Interaural Transfer Function</i>)	72
3.4.6	MWF-RTF (<i>Relative Transfer Function</i>).....	75
3.4.7	Quadro comparativo entre as técnicas apresentadas	76
4	MATERIAIS E MÉTODOS	78
4.1	GERAÇÃO DO CENÁRIO ACÚSTICO	79
4.2	AVALIAÇÃO DA FUNÇÃO CUSTO J_{ITD}	81
4.3	ANÁLISE DO ESTIMADOR DE ITD	86
4.4	HIPÓTESE DA COERÊNCIA INTERAURAL	90
4.4.1	Função custo da magnitude quadrática.....	92
4.4.2	Proposta de uma função custo integrada para preservação da ITD.....	96

5	RESULTADO EXPERIMENTAIS.....	101
5.1	CENÁRIOS ACÚSTICOS E SIMULAÇÕES	101
5.2	CENÁRIO ACÚSTICO DE CAMPO LIVRE	104
5.2.1	Experimento 1	105
5.2.1.1	<i>Avaliação Objetiva</i>	<i>105</i>
5.2.1.1.1	Erro de ITD (Δ ITD) da fonte de ruído.....	105
5.2.1.1.2	Erro de MSC (Δ MSC) da fonte de ruído	107
5.2.1.1.3	Conforto acústico.....	109
5.2.1.1.4	Qualidade do sinal de fala.....	110
5.2.1.1.5	Inteligibilidade do sinal de fala.....	112
5.2.1.2	<i>Avaliação psicoacústica</i>	<i>113</i>
5.2.1.2.1	Lateralização da fonte de ruído.....	113
5.2.1.2.2	Análise estatística dos resultados.....	115
5.2.2	Experimento 2	116
5.2.2.1	<i>Avaliação objetiva</i>	<i>117</i>
5.2.2.1.1	SNR (Conforto acústico)	118
5.2.2.1.2	Inteligibilidade e Qualidade.....	118
5.2.2.1.3	Erro de ITD (Δ ITD) e Erro de MSC (Δ MSC) da fonte de ruído.....	119
5.2.2.2	<i>Avaliação psicoacústica</i>	<i>122</i>
5.2.2.2.1	Lateralização da fonte de ruído.....	125
5.3	CENÁRIO ACÚSTICO COM REVERBERAÇÃO	130
6	CONCLUSÃO.....	134
7	PRINCIPAIS CONTRIBUIÇÕES DO TRABALHO	135
8	TRABALHOS FUTUROS	136
	REFERÊNCIAS.....	138
	APÊNDICE A – Esquema utilizado nos experimentos psicoacústicos	145
	APÊNDICE B – Artigo publicado na IEEE/ACM TASLP	148

1 INTRODUÇÃO

A habilidade que os animais e seres humanos possuem de se orientar através do sistema sensorial sempre foi fundamental para a sobrevivência dos mesmos, ajudando a escapar do predador ou na captura da presa. A orientação espacial baseada nos sentidos é um mecanismo complexo que consiste na decodificação sensorial de estímulos externos pelo sistema nervoso central. Baseado nessas informações, os seres humanos são capazes de localizar fontes sonoras no espaço e assim identificar a cena acústica na qual estão inseridos. Nesse processo, a audição e a visão são os sentidos que contribuem de maneira mais significativa, desde que trabalhando de forma coerente entre si. Por outro lado, o descasamento entre tais pistas gera sensação de desconforto nas pessoas e compromete a eficácia na localização das fontes sonoras.

Ao contrário da visão, a audição nunca é desativada completamente e por isso desempenha um importante papel de alerta (BLAUERT, 2013). Além disso, esse sentido é fundamental na vida social dos seres humanos. Por isso o presente trabalho é voltado para a lateralização¹ de fontes sonoras através da audição.

Pessoas que possuem comprometimento total na audição de uma das orelhas, situação chamada de audição monoaural, são capazes de ouvir e até mesmo lateralizar fontes sonoras com certas limitações. Entretanto, indivíduos que possuem as duas orelhas funcionais, operando com a chamada audição binaural, possuem uma série de importantes vantagens, uma vez que recebem dados adicionais sobre as fontes acústicas em um dado cenário espacial. Isto permite ao cérebro extrair informações baseadas nas diferenças entre os sinais que chegam a cada uma das orelhas. Duas dessas diferenças são usadas como pistas principais no processo de localização de fontes sonoras pontuais. São elas: a diferença de tempo interaural (ITD – em inglês *Interaural Time Difference*) e a diferença de nível interaural (ILD – em inglês *Interaural Level Difference*) (BLAUERT, 1997). Cada uma dessas pistas é gerada pela existência de dois receptores (orelhas) separados por um objeto (cabeça) com formato capaz de difratar as ondas acústicas de acordo com a posição da fonte sonora no plano horizontal (azimute). Essa difração das ondas sonoras atrasa e atenua o som que chega até a orelha mais distante da fonte sonora

¹ Nesse trabalho o termo “lateralização” refere-se à capacidade de diferenciação ou preponderância entre direita e esquerda, podendo ser associada a um ângulo de azimute VIRTUAL em relação à direção frontal de um indivíduo. Do inglês *lateralization*, “*localization of function or activity on one side of the body in preference to the other*” (Merriam-webster.com, 2017).

pontual sempre que a fonte sonora estiver posicionada em qualquer azimute diferente de zero (localização frontal ou posterior ao usuário) (BLAUERT, 1997).

Quando indivíduos perdem capacidade auditiva, aparelhos auditivos são frequentemente indicados por profissionais, como médicos e fonoaudiólogos, para compensar essas perdas. Em casos nos quais pessoas manifestam perdas em ambas orelhas, aparelhos individualmente configurados são usualmente utilizados, sendo então conhecidos como aparelhos auditivos bilaterais. Quando aparelhos bilaterais compartilham informações entre si, ou seja, cada aparelho tem também acesso aos sinais que chegam aos microfones do aparelho contralateral, eles são denominados de aparelhos auditivos binaurais (DILLON, 2001). Cada aparelho do tipo binaural geralmente conta com um conjunto de microfones que são instalados em diferentes posições, com o objetivo de aumentar o desempenho do sistema (LUTS *et al.*, 2010). Esse compartilhamento de informações proveniente de múltiplos microfones, geralmente realizado através de comunicação sem fio, amplia o acesso às características espaciais dos cenários acústicos e permite um processamento cooperativo que melhora consideravelmente o desempenho e conseqüentemente a satisfação dos usuários.

Estudos sobre a satisfação dos usuários de aparelhos auditivos indicam que conversar em ambientes contendo ruído permanece como a situação mais crítica no uso diário. Por isso, muitos trabalhos científicos têm sido endereçados ao problema de redução de ruído em sistemas binaurais, visando aumentar a qualidade e inteligibilidade do sinal de interesse (geralmente fala). Para tal, as duas principais correntes exploradas na literatura de sistemas binaurais com múltiplos microfones são baseadas em filtragem espacial (*beamforming*) (VEEN e BUCKLEY, 1988; GANNOT, BURSHEIN e WEINSTEIN, 2001; KOUTROUVELIS *et al.*, 2017) ou baseadas na filtragem Wiener Multicanal (MWF – do inglês *Multichannel Wiener Filter*) (DOCLO *et al.*, 2007; CORNELIS *et al.*, 2010; MARQUARDT *et al.*, 2015). A técnica de *beamforming* apresenta como principal desvantagem em relação ao MWF a necessidade de conhecer uma estimativa da direção de chegada da fala (DOA – do inglês *Direction Of Arrival*) para realizar a redução de ruído dentro do cenário acústico. Além disso, seu desempenho é severamente afetado quando ocorrem erros na obtenção (estimação) dessa informação de direcionalidade (GANNOT, BURSHEIN e WEINSTEIN, 2001). O MWF, por sua vez, aparece como uma opção robusta e separa o ruído da fala contaminada através do conhecimento de suas características estatísticas. Entretanto, esse tipo de filtragem pode alterar as pistas binaurais dos sinais de fala e ruído (ITD e ILD). Uma vez que as pistas binaurais de um sinal definem sua posição espacial, modificá-las significa alterar a posição percebida da fonte ou,

ainda, gerar uma percepção difusa dessa localização. Se embarcada em um aparelho auditivo binaural, essa técnica geraria confusão na identificação do cenário acústico real, podendo causar riscos às vidas de pessoas que utilizam esses aparelhos em situações cotidianas, como ao atravessar a rua conversando com alguém, andar de bicicleta, entre outros. Por esse motivo, uma ampla gama de técnicas tem sido desenvolvida visando operar em conjunto com o MWF para preservar as pistas binaurais e, por conseguinte, a lateralização correta das fontes acústicas. Essas técnicas usualmente geram uma relação de compromisso entre redução de ruído e a preservação da localização espacial da fonte de ruído. A melhor solução é aquela que apresenta uma preservação adequada às necessidades do usuário, penalizando minimamente a redução de ruído.

O objetivo desse trabalho é propor uma solução capaz de preservar a diferença de tempo interaural dos sinais processados pelo MWF, visando conseguir a correta lateralização das fontes sonoras com mínimo comprometimento da redução de sua potência.

1.1 MOTIVAÇÃO

Esta seção apresenta uma compilação de informações sob diferentes perspectivas (mercado e usuários de aparelhos auditivos) que visa demonstrar e destacar a real importância e relevância do trabalho proposto. Inicialmente foram buscados estudos atuais sobre a avaliação que usuários fazem de seus próprios aparelhos auditivos. Esses dados visam identificar e apontar as reais necessidades e dificuldades que usuários desses aparelhos enfrentam nas mais variadas situações cotidianas. Em seguida, são apresentados alguns problemas de saúde ocasionados ou associados com perdas auditivas e seu impacto em pessoas que não utilizam aparelhos auditivos. Depois são brevemente apresentadas algumas técnicas clássicas para redução de ruído aplicadas em aparelhos auditivos binaurais, assim como vantagens e desvantagens. Adicionalmente, é feita uma análise da produção acadêmica e tecnológica relativa a essas técnicas apresentadas ao longo dos tempos. Por fim, é brevemente analisado o comportamento e eficiência de umas das técnicas mais promissoras encontradas na literatura.

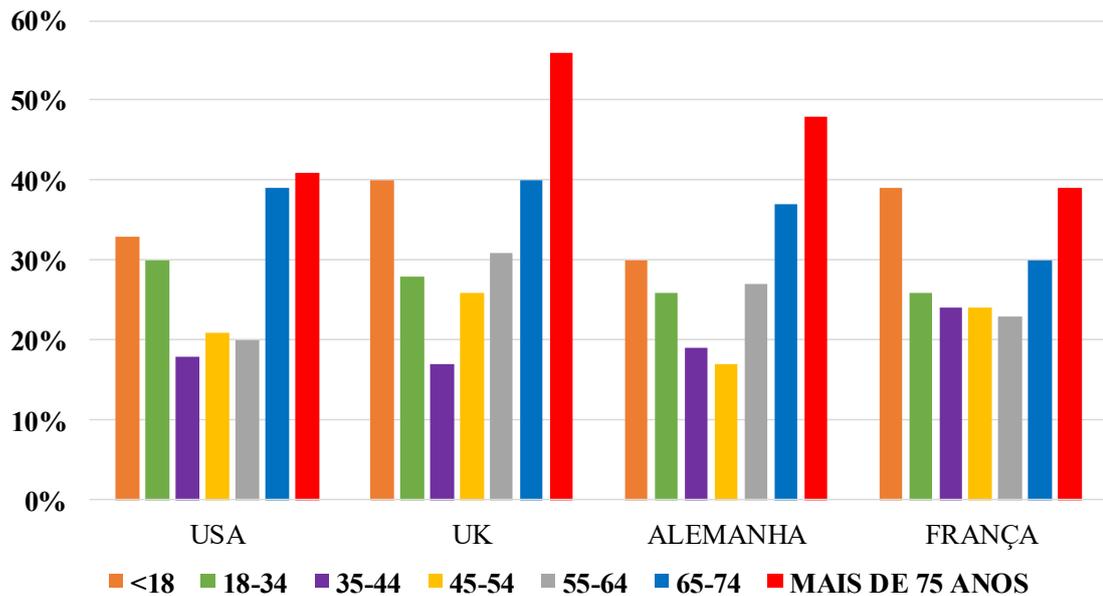
1.1.1 Tendências no mercado mundial de aparelhos auditivos

Um dos maiores desafios dos fabricantes de aparelhos auditivos é compreender o mercado atual do ponto de vista dos usuários. Além disso, uma noção de tendências é

fundamental para nortear novas ações e estratégias que os fabricantes necessitam adotar no futuro, vislumbrando consolidação e continuidade. Questões como satisfação do usuário, impactos no ambiente de trabalho, qualidade na comunicação, participação em atividades em grupo e até mesmo a qualidade de vida em geral devem ser constantemente avaliadas e servem de motivação para os fabricantes. Visando elucidar essas questões, foram criados dois grandes instrumentos de avaliação que geraram duas importantes bases de dados. Na Europa, o instrumento de avaliação é a *EuroTrak*, criado pela *European Hearing Instrument Manufactures Association* no ano de 2009 (EHIMA). Essa associação é composta por seis dos maiores fabricantes mundiais de aparelhos auditivos que são hoje responsáveis por desenvolver e produzir cerca de 90% de todos aparelhos auditivos ao redor do mundo. Nos Estados Unidos, o instrumento de avaliação é o *MarkeTrak*, desenvolvido pelo *Better Hearing Institute* no ano de 1984. Essa ferramenta conta com o apoio de 27 fabricantes de todo mundo e hoje está na sua nona versão (ROGIN, 2015). Ambos instrumentos de pesquisa são questionários aplicados *on-line* ou presencialmente em intervalos de 3 a 4 anos, e são compatíveis entre si permitindo comparações de mercados diferentes e a geração de tendências.

O primeiro resultado interessante mostrado em (EHIMA) é que, mesmo com uma estabilidade no percentual de pessoas que se autodeclararam com perda auditiva, entre os anos de 2009 a 2015, a procura por aparelhos auditivos vem aumentando em todas as faixas etárias. Como esperado, a maior adesão está no grupo de pessoas com idade superior a 75 anos, chegando a 56% na Inglaterra e 48% na Alemanha no ano de 2015, conforme a Figura 1. Isso se deve ao fato de que este grupo é o que apresenta perdas com graus mais elevados. Ao mesmo tempo, o constante aumento na expectativa de vida na maioria dos países torna crescente a quantidade de pessoas que entram nessa faixa etária e buscam qualidade de vida. Outro fator de destaque é a alta adesão de crianças e jovens, em muitos casos comparável às faixas etárias superiores, embora o percentual de pessoas autodeclaradas com perda auditiva seja baixo nessa faixa etária.

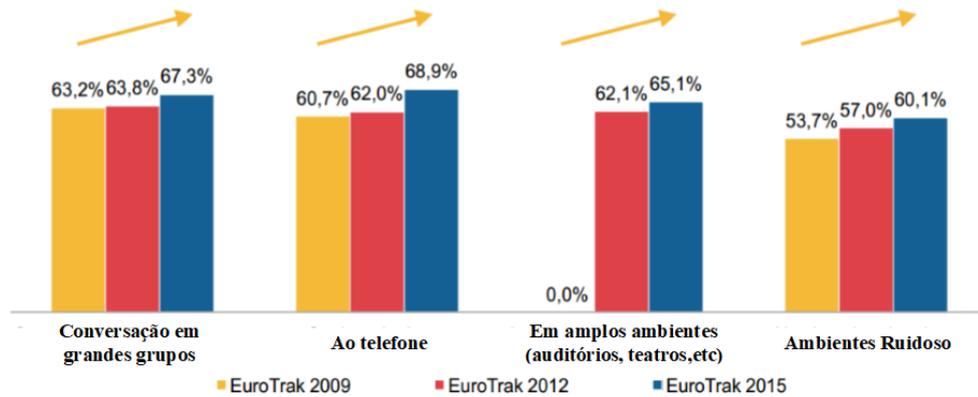
Figura 1 – Taxa de adesão ao uso de aparelhos auditivos por idade.



Fonte: (EHIMA)

A explicação para esse aumento na procura por aparelhos auditivos pode estar ligada diretamente ao aumento da satisfação dos usuários. Num passado recente, seguidamente usuários reportavam passar a maior parte do dia sem utilizar seus aparelhos ou constantemente ligando e desligando, devido aos desconfortos gerados em diversas situações cotidianas. No período em que a pesquisa foi aplicada, foi reportado um aumento de apenas 4% (de 69% para 73%) na satisfação dos usuários com relação a aspectos ligados à qualidade e conforto. Ainda que positiva, essa situação demonstra que ainda há um longo caminho a percorrer. Quando a satisfação dos usuários foi avaliada em relação ao desempenho dos aparelhos em situações consideradas críticas, como conversas em grandes grupos, uso ao telefone, uso em auditórios (reverberação) e uso em ambientes ruidosos, os resultados também foram animadores, conforme apresentado na Figura 2. Em todos os quatro cenários houve melhorias, entretanto em nenhum caso a satisfação passou de 70%. O uso de aparelhos em ambientes com ruído continua sendo o maior vilão, pois apresenta o menor percentual de satisfação (em torno de 60% em 2015). No ano de 2009, não foi avaliada a satisfação dos usuários de aparelhos auditivos em ambientes amplos.

Figura 2 – Satisfação dos usuários de aparelhos auditivos em cenas acústicas complexas.



Fonte: (EHIMA)

Outra queixa recorrente dos usuários está ligada à aparência dos aparelhos, principalmente em relação ao tamanho e forma de operar. Nesse quesito, os fabricantes foram agressivos e conseguiram uma redução de 18% no número de usuários que achavam ruim o *design* dos aparelhos. Um dos principais reflexos foi uma queda de 14% no número de pessoas que se diziam constrangidas em utilizar e operar os mesmos.

Outro dado significativo está relacionado ao aumento do número de pessoas que utilizam aparelhos nas duas orelhas (binaural/bilateral) em todos os graus de perda auditiva. Em 2009, por exemplo, 55% das pessoas usavam aparelhos bilaterais e 45% monoaurais, enquanto que em 2015 a proporção mudou para 69% e 31% respectivamente. Proprietários de aparelhos binaurais/bilaterais se consideram mais satisfeitos com relação às funcionalidades disponíveis nos aparelhos e com relação à qualidade do som nas mais variadas cenas acústicas. Por este motivo, os proprietários de aparelhos binaurais possuem um maior tempo médio de utilização diária.

Na mesma pesquisa foi mostrado que 58% dos usuários entrevistados apontam uma significativa importância dos aparelhos auditivos em suas atividades profissionais, 31% relataram importância em apenas algumas atividades e apenas 11% dos entrevistados indicaram nenhuma importância de aparelhos auditivos em suas atividades profissionais.

Em resumo, embora o número de pessoas autodeclaradas com algum grau de perda auditiva se mantenha constante, a procura por aparelhos vem aumentando nos últimos anos em praticamente todas as faixas etárias. Este aumento parece estar ligado ao aumento da satisfação dos usuários e por consequência ao tempo que o aparelho permanece ligado, impactando diretamente no ambiente de trabalho e na qualidade de vida. Usuários de aparelhos binaurais se

mostram mais satisfeitos. Entretanto, ambientes ruidosos são apontados como a pior cena acústica para a utilização de aparelhos auditivos.

1.1.2 Fatores de risco para a saúde correlacionados com perdas auditivas

Diversos trabalhos relatam que indivíduos que possuem perdas auditivas em uma ou nas duas orelhas perdem parcial ou totalmente a consciência espacial, ou seja, a capacidade para localizar as fontes que compõem a cena acústica na qual estão inseridos é alterada. Como decorrência dessas perdas, indivíduos podem ficar confusos e/ou expostos a riscos. Um exemplo é o estudo apresentado por LIN e FERRUCCI (2012), mostrando que um em cada três adultos acima de 65 anos sofre algum tipo de queda a cada ano e também que pessoas de qualquer faixa etária com perdas auditivas acima de 25 dB possuem três vezes mais chance de cair comparadas com pessoas sem perdas. Embora a causa não seja comprovada, os pesquisadores especulam que existam duas razões principais para essa conexão. A primeira está relacionada com a própria falta de consciência espacial em relação a outras pessoas, animais de estimação e objetos em geral. A segunda provável razão seria uma sobrecarga cognitiva como obstáculo ao equilíbrio de pessoas com perda auditiva. Em outras palavras, essas pessoas utilizam mais recursos mentais para ouvir, interpretar vozes e outros sons, deixando menos recursos ao sistema que mantém o equilíbrio das mesmas.

Existem também problemas de saúde mental correlacionados com perdas auditivas como a depressão, demência e Alzheimer. Em idosos, as perdas auditivas são transtornos frequentes que comprometem as atividades de socialização e conseqüentemente a qualidade de vida. Nesses casos a perda auditiva se comporta como uma doença crônica que agrava os sintomas de depressão (ACAR *et al.*, 2011). Em GURGEL, WARD, *et al.*, (2014) foi comprovado que indivíduos com perda auditiva possuem uma maior chance de desenvolver demência e Alzheimer do que indivíduos que não possuem perdas auditivas.

1.1.3 Técnicas de redução de ruído com preservação de lateralização

Os dados discutidos nas seções anteriores permitem algumas conclusões. Em primeiro lugar, os usuários de aparelhos auditivos relatam que conversas em ambientes ruidosos são as situações que acarretam maior insatisfação. Em segundo lugar, além da redução de ruído, a

preservação da localização das fontes de ruído é fundamental para dar consciência espacial ao usuário, evitando possíveis acidentes em situações cotidianas.

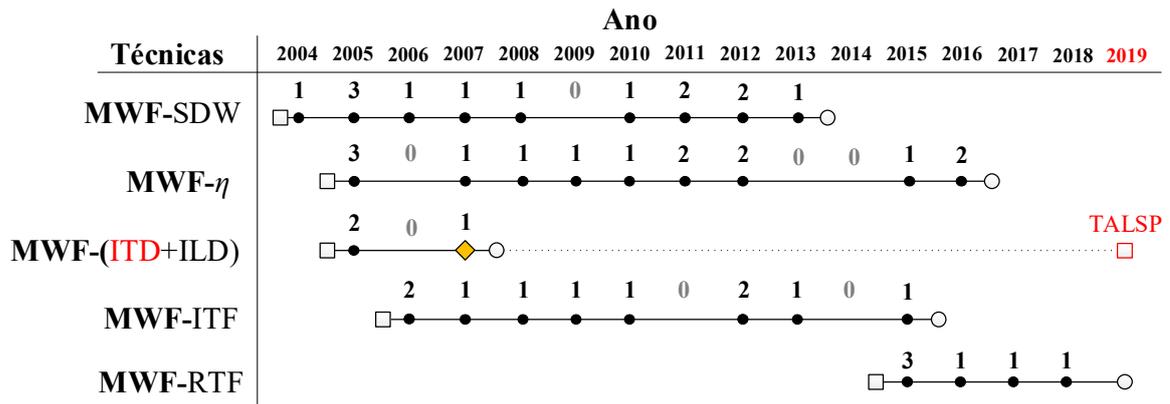
O estado da arte das técnicas de redução de ruído é encontrado em sistemas binaurais compostos por múltiplos microfones. Duas abordagens genéricas são utilizadas para esse fim. A primeira é focada na separação de todas as fontes pontuais dentro da cena acústica para então selecionar a fonte alvo (fala) baseado em uma informação adicional. A segunda usa conhecimento adicional para extrair somente a fonte alvo (fala). A segunda abordagem é mais promissora pois evita alto custo computacional para estimação de um grande número de fontes, muitas vezes irrelevantes, dentro de uma determinada cena acústica. Dentro dessa vertente existem ainda basicamente duas famílias: as baseadas em filtragem espacial e as baseadas em filtragem nos domínios tempo/frequência. Ambas as abordagens necessitam algum conhecimento prévio do sinal de interesse (fala) ou ruído. O principal desafio consiste na extração da fala enquanto todas as componentes indesejadas são suprimidas.

A técnica mais popular de filtragem espacial é conhecida como *beamforming*. Uma variação de *beamforming* amplamente encontrada na literatura aplicada a aparelhos auditivos binaurais é o MVDR (*Minimum Variance Distortionless*) (DOULO et al., 2010; JENSEN; PEDERSEN, 2015). Essa técnica visa a reduzir a potência de sinais provenientes de direções distintas da de interesse. Entretanto o principal problema atrelado ao *beamforming* é a alta sensibilidade a erros de estimação da direção de interesse ocasionados principalmente pelo descasamento dos microfones (WANG et al., 2015).

A segunda técnica mais utilizada para redução de ruído em aparelhos auditivos binaurais, é o chamado filtro de Wiener multicanal (MWF). A vantagem desse método é que ele estima os componentes do sinal de fala nos microfones de referência de cada um dos aparelhos usando a informação recebida por todos os microfones. Dessa forma, o ruído aditivo é reduzido no processo de filtragem. Além disso, essa técnica não necessita informação *a priori* sobre a configuração dos microfones nem da posição da fonte alvo (fala). Apesar dessa técnica preservar as pistas binaurais da fonte de fala, em cenas acústicas com apenas uma fonte de fala, ela modifica as pistas binaurais do ruído residual tendendo a deixá-las iguais às pistas binaurais da fala (CORNELIS et al., 2010). Dessa forma a sensação subjetiva da posição de uma fonte pontual de fala permanece inalterada enquanto que a sensação subjetiva da posição da fonte de ruído é alterada, modificando assim a cena acústica percebida. Por esse motivo, diferentes extensões da técnica de MWF foram propostas para corrigir as pistas binaurais do ruído processado e restaurar a cena acústica original.

A Figura 3 apresenta um diagrama que mostra o número de publicações associadas a cada uma das principais estratégias de preservação de pistas binaurais ao longo do tempo. As publicações elencadas na Figura 3 foram extraídas da base de dados *IEEEExplore* (IEEE Xplore Digital Library) e do portal *Web of Science* (Web of science, 2017).

Figura 3 – Ocorrência anual de publicações sobre técnicas de redução de ruído e estratégias auxiliares de preservação de localização da fonte de ruído direcional.



A primeira extensão da técnica MWF visando a preservação da localização da fonte de ruído foi denominada *Speech Distortion Weighted* (MWF-SDW) e apresentada em (SPRIET, MOONEN e WOUTERS, 2004). A ideia consiste em preservar as pistas binaurais do ruído através do aumento na distorção do sinal de fala. Vários artigos citam essa abordagem apenas como comparação com novas técnicas que foram surgindo ao longo do tempo, uma vez que ela compromete a qualidade do sinal de interesse (fala) para preservar a lateralização da fonte de ruído. Foi inicialmente apresentada em 2004, e no ano de 2013 foi encontrado um último trabalho publicado.

Logo após a introdução da primeira extensão da técnica de MWF, surgiu uma segunda estratégia denominada *Partial Noise Estimation* (MWF- η), apresentada em (KLASEN *et al.*, 2005). Nessa abordagem há um parâmetro que controla o grau da redução de ruído. A ideia se baseia em que, quanto menor for a redução de ruído, maior é a preservação natural das pistas binaurais do ruído. Quinze artigos publicados abordam e exploram esta técnica. Entretanto, suas limitações de desempenho também sempre a colocam apenas como um elemento de comparação com outras estratégias.

No mesmo ano de 2005, surgiram dois trabalhos com foco na preservação individual das diferenças de tempo e de nível interaurais da fonte de ruído (DOCLO *et al.*, 2005). Além

disso, no ano de 2007 uma patente internacional foi publicada utilizando essa estratégia (HAYKIN *et al.*, 2007) indicada com um losango amarelo na Figura 3. Nessa abordagem, a técnica que preserva a ITD é aplicada até a frequência de 1500 Hz e a técnica que preserva ILD é aplicada em frequências acima desse valor, respeitando a teoria duplex que será explicada no Capítulo 2. Embora essa abordagem fosse promissora, uma vez que respeita os princípios da psicoacústica, não foram encontrados mais artigos a respeito. Além disso, nesses dois únicos artigos não foram divulgados resultados de experimentos psicoacústicos, apenas métricas objetivas.

Em 2006, foi apresentada outra forma de abordagem que visa a preservar a função de transferência do caminho acústico entre a fonte de ruído e cada uma das orelhas, conhecida como função de transferência interaural (ITF – do inglês *Interaural Transfer Function*) (BOGAERT *et al.*, 2007). A ITF da fonte de ruído contém intrinsecamente as informações de ITD e ILD em todas as bandas de frequência. Esse é um forte indício do motivo do abandono da estratégia anterior. Foi a primeira técnica a apresentar resultados concretos de preservação da cena acústica para ruído pontual.

Por último, no ano de 2015, foi apresentada uma técnica focada na preservação da função de transferência relativa (RTF – do inglês *Relative Transfer Function*) (MARQUARDT *et al.*, 2015). A RTF é semelhante à ITF pois contém tanto informações de ITD quanto de ILD. Entretanto, sua implementação demanda a estimação das chamadas funções de transferências acústicas das fontes sonoras, o que é conhecidamente uma tarefa complexa (HUANG e BENESTY, 2003). Por esse motivo, a maioria dos trabalhos assume o conhecimento das RTFs ideais e por isso não levam em conta o impacto de eventuais erros de estimação das mesmas.

A partir do panorama apresentado, foi constatado através da literatura e de experimentos psicoacústicos que dentre todas as técnicas previamente apresentadas, as que apresentam melhores resultados são as que preservam tanto a ITD quanto a ILD simultaneamente em todas as frequências. Esse fato ignora a teoria duplex que aponta que a ITD sozinha é considerada a pista primária de localização/lateralização em frequências abaixo de 1,5 kHz e a ILD acima desse mesmo valor. Por isso, respeitar a teoria duplex acarreta na efetiva preservação da lateralização das fontes sonoras aliado a níveis superiores de redução de ruído.

1.2 PROBLEMA DE PESQUISA

Como relatado na seção anterior, uma promissora técnica cuja abordagem consiste na preservação das diferenças de tempo interaural em baixas frequências e da diferença de nível interaural em altas, de acordo com os princípios da psicoacústica, fora abandonada com apenas duas publicações no ano de 2005 e uma patente no ano de 2007. A partir dessa informação, uma bateria de experimentos objetivos foi realizada com essa solução, originando o ponto de partida do presente trabalho. Para tal, comparou-se a solução de redução de ruído baseada no método MWF convencional com a sua expansão baseada na ITD (MWF+ITD). O objetivo desses ensaios foi quantificar o impacto apenas do método baseado na ITD sobre o conforto, qualidade e inteligibilidade do áudio processado.

Tabela 1 – Avaliação da solução que visa preservar a ITD da fonte de ruído.

Técnica	Conforto	Qualidade	Inteligibilidade
Preservação da ITD	↓51,20%	↓52,12%	↓61,21%

Como pode ser observado na Tabela 1, a utilização do método baseado na ITD, diminui em mais de 50% a redução de ruído e conseqüentemente o conforto acústico do usuário do aparelho auditivo binaural. A redução no conforto impacta na diminuição do tempo de utilização diária dos aparelhos auditivos. Qualidade e inteligibilidade do sinal de fala são também severamente afetadas, dificultando a capacidade que os usuários dos aparelhos possuem de se comunicar com outras pessoas e participar de atividades em grupo. Em resumo, os resultados objetivos mostraram que essa técnica compromete os sinais de fala e ainda por cima não é capaz de preservar a localização original das fontes de ruído.

Baseado nisso um questionamento surge: “Será que com uma nova solução capaz de preservar a ITD corretamente, sendo utilizada em conjunto com a função de preservação de ILD existente na literatura (ambas respeitando os princípios da psicoacústica) seria possível obter resultados de localização tão bons quanto a solução da preservação da ITF, penalizando menos a redução de ruído?” O presente trabalho visa a trazer subsídios que contribuam para a resposta desse questionamento.

1.3 OBJETIVOS

Os objetivos do presente trabalho foram divididos em gerais e específicos.

1.3.1 Objetivos gerais

- Propor um método de redução de ruído capaz de preservar a diferença de tempo interaural (ITD) das fontes acústicas presentes em uma determinada cena acústica, visando manter a lateralidade original.
- Comparar a solução proposta com outras técnicas através de experimentos psicoacústicos e métricas objetivas e avaliar o respectivo impacto na redução de ruído.

1.3.2 Objetivos específicos

- Analisar matematicamente a solução original disponível na literatura que visa a preservação da ITD;
- Implementar a solução já existente e avaliar os resultados baseados em métricas objetivas através de simulações;
- Verificar a eficácia do método original de preservação da ITD através de experimentos psicoacústicos;
- Propor uma nova solução para a preservação da ITD;
- Avaliar o comportamento da técnica proposta através de métricas objetivas;
- Demonstrar que a nova solução é capaz de preservar a lateralização original das fontes acústicas;
- Avaliar o impacto da nova técnica em termos de redução de ruído.

1.4 ORGANIZAÇÃO DO TEXTO

O restante deste trabalho está organizado como segue. O Capítulo 2 apresenta os fundamentos para a compreensão do trabalho. Os conceitos são apresentados a partir da descrição do sistema auditivo humano e do processamento binaural usado para localização de fontes sonoras direcionais. Os principais tipos de perdas auditivas são brevemente apresentados, assim como um resumo sobre os aparelhos auditivos utilizados para compensá-las. Métricas

objetivas e subjetivas de avaliação são também apresentadas. Em sequência, no Capítulo 3, o filtro de Wiener multicanal é deduzido e aplicado como técnica de redução de ruído. As principais extensões dessa técnica visando preservação da localização da fonte de ruído são elencadas. As hipóteses para o problema de pesquisa são exploradas no Capítulo 4, justificando assim a proposição de uma nova abordagem para o problema. No Capítulo 5 são apresentados resultados experimentais de avaliação da nova abordagem. No Capítulo 6, são apresentadas as conclusões do estudo aqui apresentado, enquanto que nos Capítulos 7 e 8 são apresentadas, respectivamente, as principais contribuições do trabalho e propostas de continuação.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo alguns conceitos importantes para a compreensão desse trabalho serão apresentados. Inicialmente é feita uma breve apresentação dos elementos que compõem o sistema auditivo dos seres humanos, assim como as diferentes formas em que os sons são transmitidos da orelha externa até a orelha interna. Em seguida, é apresentada uma discussão de como o cérebro decodifica as informações provenientes da cóclea, e quais pistas são utilizadas para a localização espacial de fontes sonoras.

Quando um indivíduo tem sua capacidade auditiva reduzida, pelos mais variados motivos, diz-se que apresenta perdas auditivas. Além da redução da capacidade de compreensão da fala, as perdas auditivas também reduzem a capacidade de localizar fontes sonoras pontuais. Por esse motivo, um breve resumo sobre os diferentes tipos de perdas e os diferentes graus em que as mesmas são classificadas é apresentado.

Um das soluções mais utilizadas para a correção dessas perdas é o aparelho auditivo. Esses dispositivos são amplificadores de áudio, configurados para compensar as perdas individuais. Devido ao fato que o presente trabalho propõe soluções que serão embarcadas nesses dispositivos, uma sucinta explicação sobre os mesmos é apresentada.

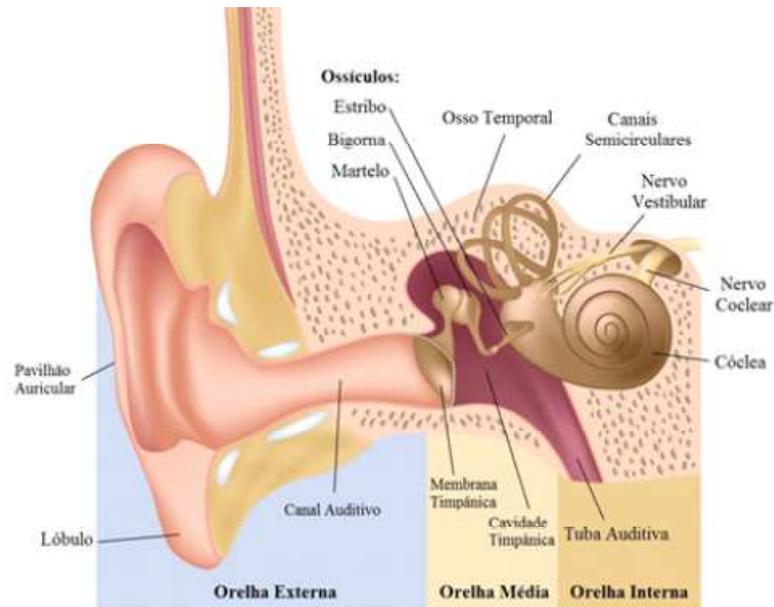
Por fim, são também apresentadas métricas para avaliação da fala.

2.1 ANATOMIA DO SISTEMA AUDITIVO HUMANO

O objetivo desta seção é apresentar os elementos principais que compõem o sistema de audição humano e o papel que cada um desses elementos desempenha. Para a melhor compreensão desse sistema vamos basear a explicação na representação mostrada na Figura 4. Como podemos observar na mesma figura, o sistema auditivo humano é dividido em três partes principais, de acordo com a estrutura, modo de operação e a função que desempenha (TATE, 1994).

A orelha externa é composta pelo pavilhão auricular e o canal auditivo externo cuja função é a captação e condução do som até a orelha média. A orelha média, por sua vez, é um sistema composto pela membrana timpânica e três ossículos denominados martelo, bigorna e estribo. Sua função é transmitir as vibrações da membrana através dessa cadeia de ossículos até a orelha interna. Essa transmissão pode ocorrer de três maneiras distintas:

Figura 4 – Representação da anatomia do sistema auditivo humano.



Fonte: (TEXAS EAR CENTER, 2017).

1. Por condução óssea;
2. Por mudanças na pressão do ar na cavidade da orelha média;
3. Através da cadeia de ossículos da cavidade da orelha média até a orelha interna (método natural).

No método natural de audição, o movimento do estribo comprime o fluido e outras estruturas da orelha interna. Esses estímulos se propagam até a cóclea e são captados pelas células ciliadas, que por sua vez convertem tais estímulos em impulsos elétricos que são transmitidos através do nervo auditivo. Tal processo é conhecido como transdução mecânico-neural. Além disso, a orelha interna é responsável por abastecer o sistema nervoso com informações sobre frequência, intensidade e o conteúdo temporal da estimulação acústica. Parte da composição espectral do som é provida pela mecânica da orelha interna que pode ser descrita como um processo de filtragem (BORISAGAR, THANKI e SEDANI, 2019). As estruturas de canais circulares e vestíbulos afetam apenas o equilíbrio ao invés da audição, por isso não serão abordados no trabalho.

2.2 PROCESSAMENTO BINAURAL DE FONTES SONORAS DIRECIONAIS

Como visto na seção anterior, os sons são transmitidos através de uma cadeia complexa até serem convertidos em informação neural pela orelha interna. Nesta seção é apresentada a forma pela qual o cérebro usa essas informações binaurais para mapear a localização de uma fonte sonora pontual. Essa capacidade de localização de fontes sonoras permite aos seres humanos obter uma atenção seletiva para fontes acústicas de interesse em detrimento de outras. Esse fenômeno pode ser exemplificado pelo chamado efeito “*cocktail party*”, no qual um indivíduo tenta estabelecer um diálogo em uma festa com música, ruído e outras pessoas conversando por toda parte. Para estabelecer o diálogo e compreender o conteúdo da conversa, o indivíduo deve focar sua atenção auditiva na fala do outro indivíduo envolvido no diálogo, ao mesmo tempo que coloca as demais fontes sonoras em segundo plano. A capacidade de separar a fonte de interesse e enfatizá-la melhora o entendimento do diálogo.

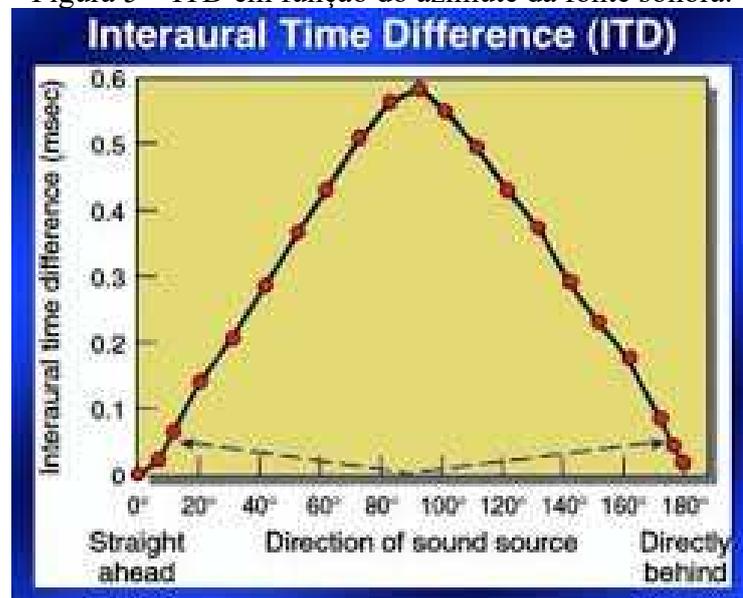
A localização de fontes sonoras é possibilitada por pistas acústicas extraídas das diferenças do campo sonoro em cada uma das orelhas (CARLILE, 2013). Obviamente, tais diferenças são geradas por modificações que sofrem os sinais que chegam até as orelhas em função da localização da fonte. Essas modificações são constituídas basicamente por atrasos e atenuações gerados pelo tamanho da cabeça humana que reflete e refrata ondas sonoras, efeito conhecido como *head shadow*. Quando o som trafega de uma fonte localizada diretamente a frente ou atrás de um indivíduo, ele irá atingir as duas orelhas no mesmo instante de tempo e com a mesma intensidade. Em qualquer outro caso, quando a fonte sonora estiver deslocada para um lado do plano sagital médio do indivíduo, o som chegará até a orelha mais distante da fonte sonora com (i) um atraso em relação à orelha mais próxima, introduzindo uma diferença de tempo interaural, e (ii) com uma intensidade menor em relação à orelha mais próxima, introduzindo uma diferença de nível interaural.

2.2.1 Relação entre azimute e as pistas ITD e ILD

Como mencionado anteriormente, a diferença de tempo interaural é um efeito gerado pela posição da fonte em relação às orelhas que defasa os sinais que chegam a cada uma das orelhas. A máxima ITD para seres humanos, considerando a velocidade média do som (~340 m/s) e a distância média interaural de um adulto (22 - 23 cm), é por volta de 670 μ s, ocorrendo quando a fonte estiver diretamente à esquerda ou à direita (posições associadas com

os azimutes de $\pm 90^\circ$, respectivamente) (BLAUERT, 1997). Na Figura 5, é apresentada a relação da ITD em milissegundos com o azimute de uma fonte sonora. Como mencionado anteriormente, o azimute de $+90^\circ$ apresenta a maior ITD enquanto que o azimute de 0° está associado a um valor nulo de ITD. Outra observação importante na Figura 5 é que cada valor de ITD aponta para dois azimutes diferentes (linha azul tracejada). Um indica que a fonte está atrás da linha média da cabeça (posterior) e o outro na frente dessa linha (anterior ou frontal). Essas ambiguidades geram o erro de localização/lateralização conhecido por “*front/back confusion*”.

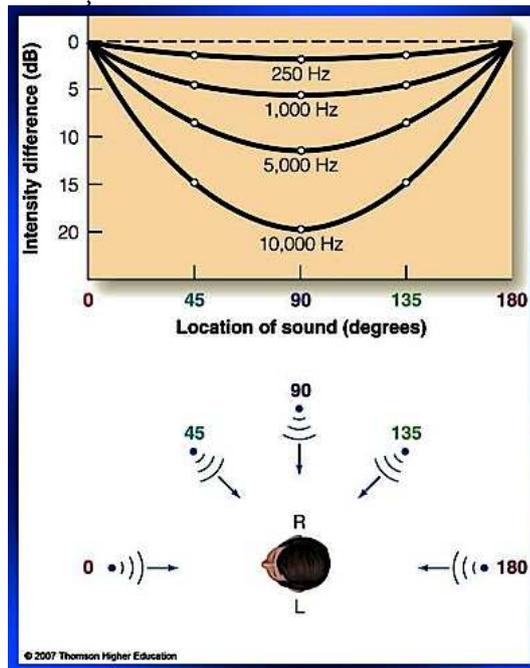
Figura 5 – ITD em função do azimute da fonte sonora.



Fonte: (STAAB, 2015).

Da mesma forma que a cabeça atrasa o som que chega até a orelha mais distante da fonte de ruído, ela também atenua esse som, gerando uma ILD. A Figura 6 mostra como a ILD varia em função do azimute da fonte para diferentes frequências. É possível observar que as variações são maiores à medida que a frequência aumenta. Frentes de onda de baixas frequências (250 Hz) geram pequenas ILDs enquanto que frentes de onda de altas frequências (10 kHz) geram ILDs muito maiores.

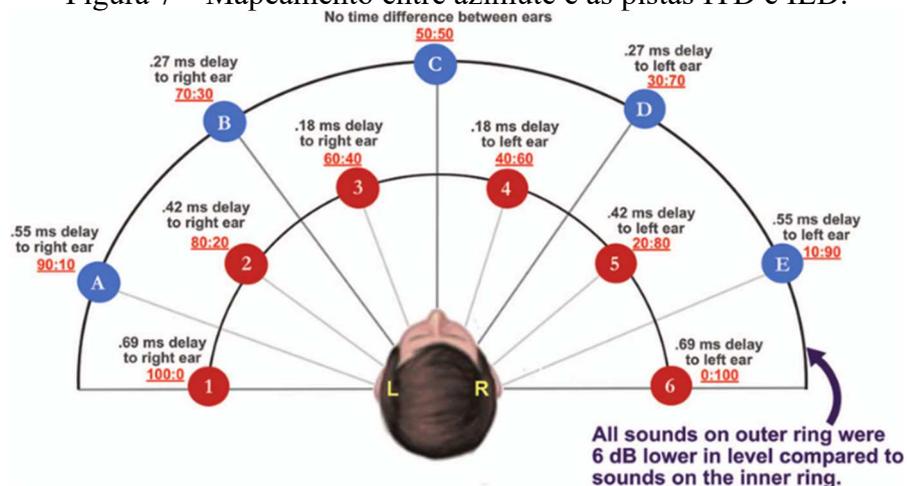
Figura 6 – Variação da ILD com o azimute da fonte direcional.



Fonte: (STAAB, 2015).

A Figura 7 apresenta uma espécie de mapa de localização em que cada ponto marcado com letras e números equivale a um azimute diferente. Atrélado a cada ponto estão marcados o valor da ITD em milissegundos (em azul) e embaixo a ILD proporcional (percentual da orelha esquerda: percentual da orelha direita) em vermelho. Os pontos do semicírculo interno apresentam níveis maiores do que os pontos do anel externo devido à menor distância entre a fonte sonora e o indivíduo.

Figura 7 – Mapeamento entre azimute e as pistas ITD e ILD.



Fonte: (FAIRNIE, MOORE e REMINGTON, 2016).

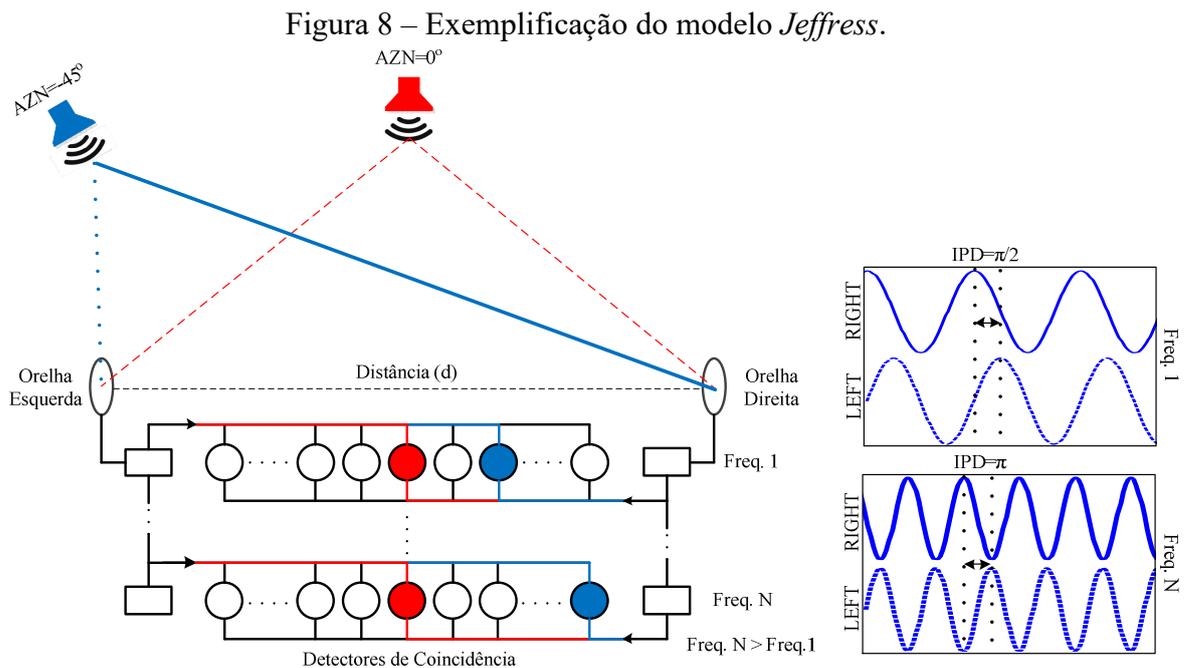
2.2.2 Teoria Duplex

A teoria duplex descreve o processo de localização de sons através de pistas extraídas das diferenças do campo acústico entre cada uma das orelhas, tendo sido enunciada pela primeira vez no trabalho de Lord Rayleigh no ano de 1907 (RAYLEIGH, 1907). Embora nesse estudo o comportamento das ondas sonoras em volta da cabeça tenha sido modelado com aproximações simplificadas (a cabeça como uma esfera e as orelhas como dois receptores pontuais simetricamente posicionados), a teoria duplex tem influenciado e dominado as pesquisas da área durante o último século. Essa teoria sustenta que a ILD causada pelo efeito conhecido como *head shadow* é a principal pista binaural usada pelos humanos para localização de sons de alta frequência (acima de 1,5 kHz) no plano azimutal. Abaixo dessa frequência, essas diferenças de intensidade são muito menores e por isso a localização de sons é dominada pela ITD (SMITH e PRICE, 2014).

Embora muito utilizada, a teoria duplex é incompleta visto que existe um grande número de observações que não podem ser explicadas com esta teoria, como também outro grande número de experimentos que inclusive contrariam as suas premissas. O primeiro ponto de divergência está ligado à delimitação das bandas controladas pela ITD e ILD. O segundo consiste no peso de cada pista, não dominante, dentro de cada banda. Existem experimentos baseados na localização de tons puros e na localização de sinais de banda larga. Nos experimentos com tons puros, Lord Rayleigh iniciou afirmando que 500 Hz seria o limite entre as duas informações, sendo que a ITD dominaria em frequências abaixo de 128 Hz e que a ILD dominaria acima de 500 Hz. Dois anos depois, Lord Rayleigh expressou que a ITD não poderia ser usada acima de 400 Hz (STRUTT, 1909). Desde então, diversos trabalhos tentaram identificar esse limite para tons puros. Em STEVENS e NEWMAN (1936), o limite de informação útil provida pela ITD foi definido abaixo de 2 kHz e para informação de ILD acima de 4 kHz enquanto que MILLS (1960) e SANDEL, TEAS, *et al.* (1955) definiram o limite único de 1,5 kHz, que será utilizado no presente trabalho. MACPHERSON e MIDDLEBROOKS (2002) mostraram que a teoria duplex também se aplica a sons de banda estreita, nos quais a ITD tem um peso maior em frequências baixas e a ILD possui um peso maior em frequências altas. Em experimentos de localização com fones de ouvido desenvolvidos por WIGHTMAN e KISTLER (1992), para simular experimentos de localização de ruído de banda larga em campo livre, a posição da fonte foi determinada pela ITD sempre que o ruído incluía potência abaixo de 2,5 kHz.

2.2.3 Modelo *Jeffress* para ITD

O modelo de *Jeffress* foi proposto pelo psicólogo *Lloyd Jeffres* no ano de 1948 (JEFFRESS, 1948), na tentativa de explicar como os animais conseguem identificar a direção de chegada de fontes sonoras através da ITD. O autor descreve o sistema binaural como duas linhas de atraso, cada uma conectada a uma das orelhas. Essas linhas de atraso são compostas por neurônios detectores de coincidência binaurais que respondem a estímulos sonoros. A ideia é que diferentes detectores de coincidência são ativados para diferentes ITDs. A Figura 8 exemplifica o processo mostrando uma cena acústica composta por duas fontes sonoras direcionais, uma no azimute de 0° (vermelha) e a outra no azimute de -45° (azul). A fonte vermelha encontra-se à mesma distância das duas orelhas, por isso em todas as frequências o tempo de chegada nas duas orelhas é o mesmo. Dessa forma, em todas as frequências o detector de coincidência central é ativado. No caso da fonte de azimute -45° , a fonte sonora está mais próxima da orelha esquerda do que da orelha direita do indivíduo, fazendo com que diferentes detectores de coincidência sejam ativados de acordo com a frequência. Esses detectores de coincidência geram diferenças de fase para cada frequência, conhecidas como atrasos de fase interaural (IPD – do inglês *Interaural Phase Delay*). Quanto maior a frequência do sinal sonoro, maior a IPD gerada como mostra o exemplo da Figura 8 com duas frequências.



A preservação das IPDs em suas respectivas frequências acarreta na preservação das características temporais da diferença de tempo interaural de estrutura fina (TFS – do inglês *Temporal Fine Structure*), que correspondem às rápidas variações com taxas próximas da frequência central da banda. A ITD da estrutura fina é considerada a pista dominante para localização de fontes audíveis de frequência inferior a 1,5 kHz (WIGHTMAN e KISTLER, 1992). O presente trabalho visa a preservação das pistas binaurais de sinais de frequência inferior a 1,5 kHz, por esse motivo a ITD da estrutura fina é a pista que se deseja preservar.

Sinais com frequências mais altas (acima de 1,5 kHz) apresentam comprimentos de onda menores do que a distância entre as orelhas, por conta disso, o *head shadow* é gerado e a pista de ILD se torna responsável pelo processo de localização. Dessa forma, essas frequências não contribuem para a localização através da ITD da estrutura fina. Para mamíferos com cabeça menor, como por exemplo em ratos, é possível detectar ITDs através de frequências mais elevadas. Para sinais com frequência acima de 1,5 kHz o chamado de ITD do envelope temporal (TE – do inglês *Temporal Envelope*), que corresponde às lentas variações em amplitude dos sinais no domínio do tempo, é a característica responsável junto com a ILD pela localização de fontes sonoras pontuais.

2.2.4 Limite anatômico imposto pelo tamanho da cabeça humana no uso da ITD

Como já discutido no Capítulo 2, a teoria duplex estabeleceu que o cérebro utiliza a ITD como pista dominante para localização de sinais com frequência abaixo de 1,5 kHz e a ILD como pista dominante para sinais acima desse limite. Atualmente, ainda há muita discussão sobre o exato valor de frequência que separa a utilização de cada uma das pistas, assim como qual é a contribuição da pista não dominante em cada banda no processo de localização.

Em HARTMANN e MACAULAY (2014) foi apresentada uma nova interpretação sobre esse tema que questiona se realmente a ITD é a única pista usada para a lateralização de fontes sonoras nas frequências abaixo de 1,5 kHz. Nessa análise, os autores utilizam o modelo esférico de cabeça para calcular a ITD gerada em diferentes azimutes entre 0° e $\pm 90^\circ$, através da seguinte fórmula:

$$ITD = \left(\frac{3a}{v} \right) \sin(\theta) \quad (1)$$

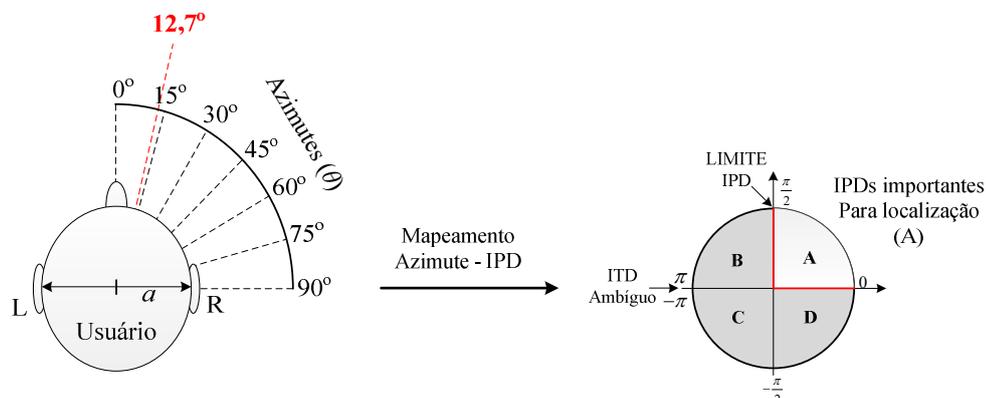
em que a é o raio médio da cabeça humana (8,75 cm), v é a velocidade do som na temperatura ambiente (340 m/s) e θ (em graus) corresponde ao azimute da fonte sonora. Esse modelo não considera a distância da fonte, uma vez que existe pouca variação da ITD relacionada à distância, ao contrário da ILD que é muito mais sensível às variações da distância da fonte até a cabeça. Segundo esse modelo, o maior valor de ITD é de $763 \mu\text{s}$ quando a fonte pontual está em campo aberto no azimute $\theta = 90^\circ$ como mostrado na Tabela 2.

A Figura 9 apresenta um diagrama que representa o mapeamento entre o azimute da fonte pontual (apenas valores positivos) para uma diferença de fase interaural (IPD), para cada frequência, calculada através de:

$$\text{IPD} = \text{ITD} \cdot \omega \quad (2)$$

As IPDs são separadas em quatro quadrantes (**A**, **B**, **C** e **D**) que impactam diretamente na capacidade de localização. Os quadrantes **C** e **D** são considerados ambíguos e responsáveis por gerar imagens da fonte pontual no lado contrário da posição real da fonte. Esse efeito indesejado ocorre quando a IPD ultrapassa 180° e foi documentado em experimentos com fones de ouvido em BERNSTEIN e TRAHOTIS (1985). Em SAYERS (1964) foi mostrado que IPDs na região **B** podem gerar a formação de uma imagem atrás do plano frontal assim como também lateralizar imagens no lado contrário da cabeça (plano sagital). YOST (1981) também encontrou casos de lateralização em hemisfério contrário à posição original com IPDs na região **B**. Sendo assim, conclui-se que apenas as IPDs do quadrante **A** contêm informação confiável para a correta lateralização de fontes sonoras.

Figura 9 – Mapeamento de um azimute para uma IPD.



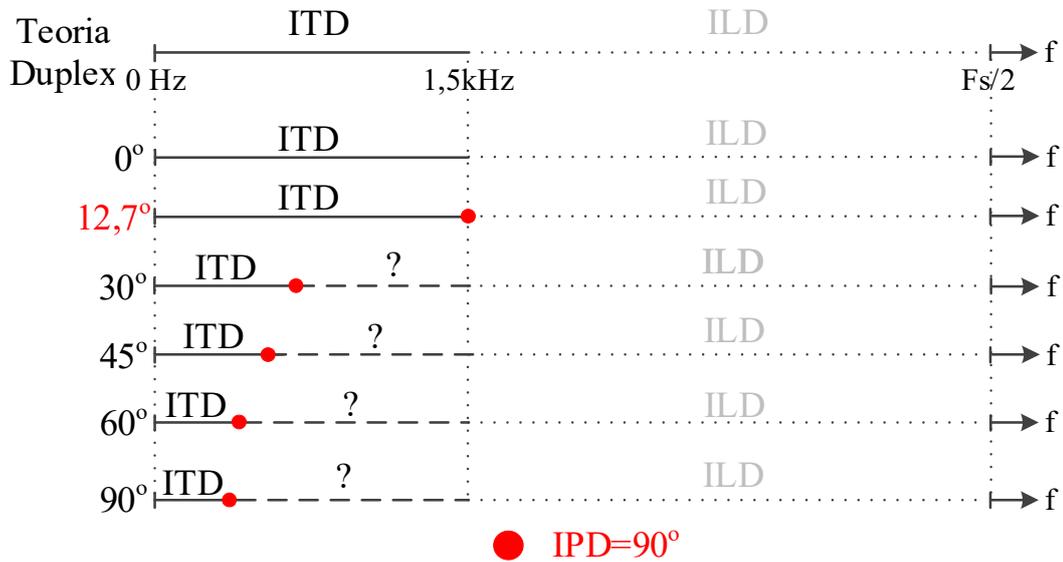
Ainda na Figura 9, observamos que o azimute de $12,7^\circ$ encontra-se em destaque pois, segundo o modelo esférico da cabeça, é o maior azimute capaz de mapear totalmente a banda de 1,5 kHz em IPDs dentro do quadrante A. À medida que a fonte de ruído ultrapassa esse valor de azimute de $12,7^\circ$, uma banda de frequências acaba tendo IPDs fora do quadrante A (banda crítica), podendo interferir equivocadamente na lateralização das fontes. Esse problema pode ser analisado com mais detalhes através da Tabela 2 (coluna 3), onde é possível verificar que existe uma frequência limite para a informação útil de IPD de acordo com o azimute. Também é possível constatar que à medida que o azimute aumenta, maior é a banda (banda crítica) que cai dentro de quadrantes não confiáveis. No caso do azimute de $+90^\circ$, por exemplo, apenas por volta de um quinto da banda de 1,5 kHz (327,65 Hz) apresenta IPDs no quadrante A, enquanto que todo o resto (banda crítica) se encontra em outros quadrantes.

Tabela 2 – Mapeamento de azimute para ITD, banda útil e banda crítica para localização de fontes direcionais baseados no tamanho médio da cabeça humana.

Azimute (θ)	ITD	Limite IPD Banda Útil (BU)	IPD > $\pi/2$ (B, C, D) Banda Crítica
0°	0	-	-
$+12,7^\circ$	167,7 μ s	1500 Hz	0
$+15^\circ$	197,47 μ s	1266 Hz	234 Hz
$+30^\circ$	381,54 μ s	655,24 Hz	844,76 Hz
$+45^\circ$	539,52 μ s	463,37 Hz	1036,63 Hz
$+60^\circ$	660,77 μ s	378,34 Hz	1121,66 Hz
$+75^\circ$	737 μ s	339,21 Hz	1160,79 Hz
$+90^\circ$	763 μ s	327,65 Hz	1172,35 Hz

Na prática, o que acontece é o surgimento de uma terceira banda (banda crítica), representada na Figura 10 pelo sinal de interrogação. Nessa banda não há um consenso sobre como e quais pistas são realmente utilizadas pelo cérebro no processo de lateralização. Talvez o raciocínio simplista da teoria duplex que em baixas frequências se utilize apenas ITD e acima apenas ILD não seja binário, e sim algo cooperativo entre as duas pistas. É, talvez, por esse motivo que as técnicas que preservam a ITF do ruído apresentam resultados tão bons de lateralização, uma vez que a ITF preserva ambas as pistas simultaneamente.

Figura 10 – Modificação da teoria duplex segundo a interpretação ecológica.



Como resultado, o estudo em HARTMANN e MACAULAY (2014) conclui que a explicação ecológica seria válida apenas se a cabeça humana fosse em torno de 50% menor. Talvez no processo de evolução dos seres humanos, no qual a cabeça teve seu tamanho constantemente aumentado ao passar dos anos, o cérebro tenha desenvolvido outros mecanismos mais complexos de localização dentro da banda crítica além da ITD. Embora importantes, essas considerações acerca dos limites anatômicos não serão utilizadas nos experimentos apresentados no trabalho. Entretanto, acredita-se que as mesmas possam ser úteis para proposições futuras de novas estratégias.

2.2.5 Peso da ILD na banda dominada pela ITD

Na seção anterior, foram discutidos aspectos relacionados à anatomia da cabeça humana, e que seu respectivo tamanho médio é responsável por gerar valores de ITDs que são mapeadas em IPDs conflitantes no processo de localização de fontes sonoras pontuais. A partir daí, foi mostrado que para cada azimuth existe uma frequência que divide a banda comandada pela ITD em banda útil e banda crítica. Nessa seção será investigado se a ILD é capaz de contribuir de alguma forma dentro da banda crítica para resolver os conflitos gerados.

Em HARTMANN, RAKERD e CRAWFORD (2016) foi publicado um estudo que avalia a influência da ILD no processo de lateralização de tons senoidais com frequências de

250 Hz, 500 Hz e 750 Hz em campo livre (sem fones de ouvido). Foram selecionadas cinco pessoas para uma bateria de experimentos subjetivos de localização em quatro situações distintas:

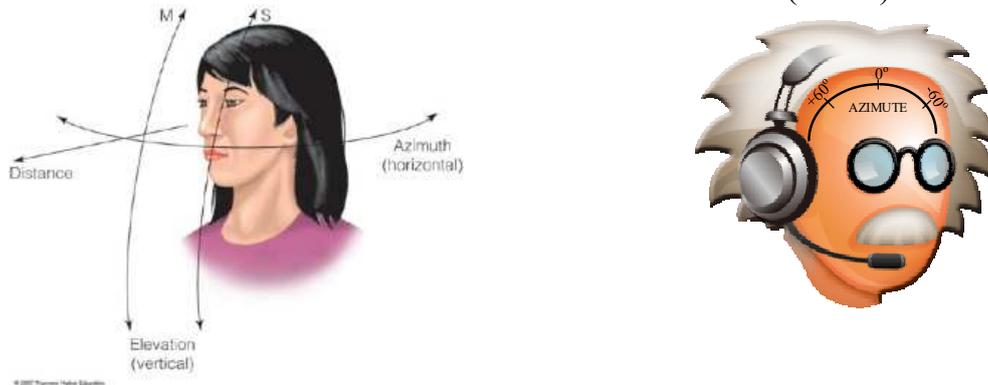
- i. ILD naturalmente coerente com o azimute da fonte pontual;
- ii. $ILD = 0$;
- iii. ILD fixo com 6 dB em discordância com a ITD;
- iv. ILD fixo com 12 dB em discordância com a ITD.

A primeira constatação é que quando a ILD é coerente com a ITD (Situação i) há uma modesta melhoria na percepção correta do azimute para tons em todos os valores de ITD e a eliminação das confusões de localização (direita-esquerda) que ocorrem quando a IPD ultrapassa 180° . Aumentar a ILD para valores improváveis de mesmo sinal (hemisfério) pode piorar a percepção de lateralidade, aumentando a confusão *front-back*. Em experimentos com ILD igual a zero (Situação ii) também foram reportados erros de inversão de hemisfério, principalmente para a frequência de 750 Hz com azimutes acima de 60° . Nos experimentos onde a ILD discorda em sinal (hemisfério) da ITD (Situações iii e iv), foi mostrado um efeito chamado ITD *slipped-cycle*, que também pode gerar erros de inversão de hemisfério. Como conclusão, o trabalho afirma que a ILD é capaz de inverter o hemisfério de sinais acústicos com frequências pertencentes à banda da ITD, demonstrando que a teoria duplex não é uma regra que estabelece que apenas uma pista comanda individualmente cada banda.

2.2.6 Localização versus lateralização de fontes sonoras direcionais

Uma dúvida recorrente é relacionada às diferenças conceituais de localização e lateralização de fontes sonoras direcionais. A localização é a habilidade de determinar a posição de uma fonte sonora em um campo aberto (sem fones de ouvido). Lateralização, por outro lado, é relacionada à posição virtual do som percebido dentro da cabeça para estímulos aplicados com fones de ouvido (PLENGE, 1974). A Figura 11 apresenta a representação de dois indivíduos tentando localizar (esquerda) e lateralizar (direita) uma fonte sonora direcional.

Figura 11 – Indivíduo localizando fonte sonora em campo aberto (esquerda) e lateralizando fonte sonora com o auxílio de fones de ouvido (direita).



O processo de localização de uma fonte é mais complexo e permite o julgamento da direção aparente (azimute e elevação) e da distância desta fonte em campo aberto. A localização se baseia na integração sensorial de uma ampla gama de pistas e ainda não há um consenso do peso de cada uma delas nesse processo. Experimentos práticos de localização são mais completos, porém são custosos, visto que demandam uma considerável infraestrutura como sistemas de som e câmaras anecoicas ou reverberantes.

Experimentos de lateralização são na prática muito menos custosos, pois demandam apenas um computador e um par de fones de ouvido. Os efeitos reais de campo aberto relacionados ao ambiente, assim como o azimute, são emulados através de funções de transferência relacionadas à cabeça (HRTF – do inglês *Head-Related Transfer Function*) que mapeiam o caminho acústico entre a fonte sonora e cada uma das orelhas do ouvinte. Essas funções de transferência são obtidas em experimentos em campo aberto com a utilização de um torso e cabeça de um manequim. Na prática, essas HRTFs modificam os sons que chegam a cada uma das orelhas dando ao usuário de fones de ouvido a sensação subjetiva associada a uma cena acústica determinada.

2.3 GRAUS E TIPOS DE PERDAS AUDITIVAS

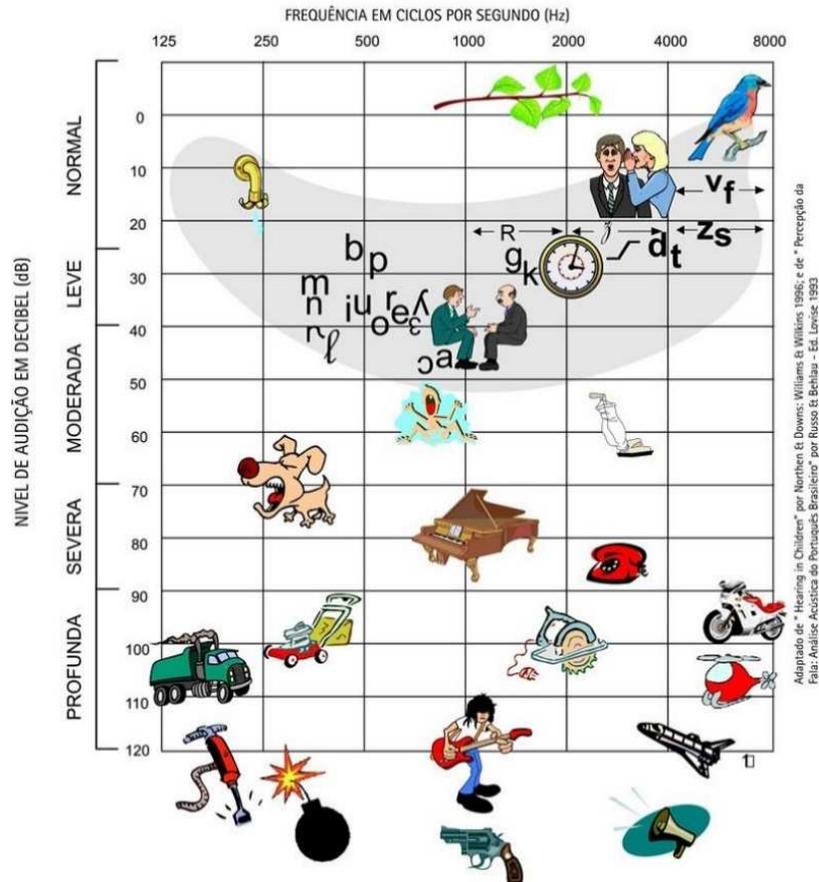
O problema que estamos tentando resolver neste trabalho surge devido à necessidade do uso de aparelhos auditivos para compensar as chamadas perdas auditivas. As características da perda auditiva em um indivíduo dependem do tipo e da gravidade do problema que a causou. Uma forma muito utilizada para avaliar o grau e o tipo de perda auditiva é o exame audiométrico, aplicado por médicos e fonoaudiólogos. Nesse exame, o profissional classifica o

grau da perda auditiva através de uma escala, como por exemplo a apresentada por DAVIS e SILVERMAN (1970). Segundo essa escala, uma pessoa normal tem, por exemplo, a capacidade de ouvir sons de intensidade inferior a 25 dB de nível de pressão sonora (SPL – do inglês *Sound Pressure Level*). Caso não consiga, o indivíduo é considerado com perda auditiva. As perdas são classificadas conforme a intensidade sonora mínima que um indivíduo é capaz de escutar e são divididas conforme os seguintes graus:

- **Leve (26 a 40 dB SPL):** as pessoas geralmente não percebem que estão ouvindo menos e tendem a aumentar progressivamente a intensidade da voz. Nesse grau, os sons de menor intensidade são cada vez mais difíceis de serem escutados e conversar em ambientes ruidosos torna-se uma tarefa difícil.
- **Moderada (41 a 70 dB SPL):** quando a perda é moderada quase nenhum som de fala pode ser percebido em nível de voz natural. Geralmente é nesse grau que as pessoas começam a buscar ajuda de médicos e fonoaudiólogos visando a sanar o problema.
- **Severa (71 a 90 dB SPL):** quando a perda é severa, nenhum som de fala se torna audível em nível normal de intensidade de voz e conversas em grupos se tornam praticamente impossíveis.
- **Profunda (acima de 91 dB SPL):** a perda auditiva profunda impede que as pessoas escutem a maioria dos sons, percebendo apenas sons graves que transmitem vibração como um avião, trovão, etc.

A Figura 12 mostra um diagrama que resume cada um dos graus de perda auditiva em termos de frequência e intensidade segundo DAVIS e SILVERMAN (1970). Além disso, alguns dispositivos como aspirador de pó, motocicleta, piano, entre outros, são posicionados em relação à intensidade e frequência que eles emanam sonoramente.

Figura 12 – Relação entre os diferentes graus de perdas auditivas com sons encontrados no cotidiano.



Fonte: (AUDIOMAX, 2016).

A área em cinza mais escuro na Figura 12 corresponde à intensidade e frequência da fala, mostrando que mesmo perdas leves já são capazes de comprometer a compreensão de uma simples conversa. Na prática, pessoas com perdas auditivas não conseguem ouvir alguns tipos de sons. Pessoas com perdas severas ou profundas podem não conseguir ouvir qualquer som de fala, a não ser que sejam pronunciados em uma intensidade muito maior que a normal e a uma curta distância. Indivíduos com perdas moderadas são capazes de ouvir alguns determinados tipos de sons enquanto outros não. Em particular, alguns fonemas, geralmente as consoantes, podem simplesmente não ser escutados. Para tornar esses sons audíveis, aparelhos auditivos fornecem amplificação seletiva visando a compensar as perdas e corrigir o problema.

Além do grau da perda auditiva, os profissionais devem determinar também através da audiometria o tipo de perda auditiva que o indivíduo apresenta. Esse diagnóstico é muito importante e pode indicar o tipo de tratamento que será necessário. Existem três principais tipos

de perda auditiva, a condutiva, a neurossensorial e a mista, que serão apresentadas a seguir (MALTBY, 2005).

Uma perda auditiva condutiva se deve a problemas na orelha externa e média que impedem que o som chegue efetivamente até a orelha interna. Por exemplo, um bloqueio ou má-formação na orelha média são capazes de impedir que os ossículos (martelo, bigorna e estribo) vibrem de maneira correta. As causas mais comuns desse tipo de perda são infecções na orelha média ou ruptura da membrana timpânica. Geralmente a perda condutiva é temporária e pode ser tratada com medicamentos ou cirurgia. Entretanto, existem algumas perdas condutivas que não podem ser tratadas, esses casos são chamados de perdas condutivas permanentes. Nessas situações, podem-se utilizar implantes de orelha média ou aparelhos auditivos convencionais.

A perda auditiva neurossensorial é o resultado de um problema na orelha interna, mais precisamente quando as células ciliadas da cóclea estão ausentes ou danificadas, por isso esse tipo de perda é geralmente permanente. As células ciliadas são responsáveis pela produção de sinais elétricos que o cérebro utiliza para interpretar corretamente o som. É também conhecida como “surdez neural” e é o tipo de perda auditiva mais comum.

Algumas pessoas apresentam perda auditiva apenas em altas frequências, a isto se chama surdez parcial. Nesses casos apenas as células ciliadas da base da cóclea estão danificadas enquanto que as células da parte interna da cóclea, responsáveis pela captação das baixas frequências, permanecem intactas.

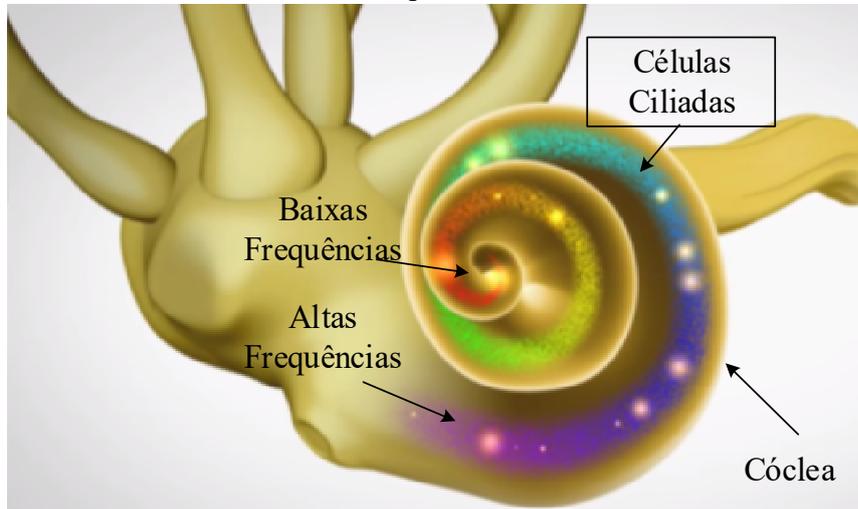
Existem diversas causas para este tipo de perda, como:

- Hereditariedade/genética;
- Perda auditiva ocasionada por excesso de ruído;
- Lesões na cabeça;
- Determinadas medicações que causam danos às células ciliadas;
- Doenças como sarampo, caxumba, meningite;
- Envelhecimento;
- Congênitas;
- Tumores no nervo auditivo.

Na Figura 13 é mostrada uma ilustração da cóclea, sendo colocadas em evidência (colorido) as células ciliadas. Vale ressaltar que as células localizadas na base da cóclea são responsáveis por converter em sinais elétricos as componentes de alta frequência dos sinais

sonoros, enquanto que na parte central estão as células que convertem os sinais de baixa frequência.

Figura 13 – Ilustração da cóclea com ênfase nas células ciliadas separadas por bandas de frequência.



Fonte: (MED-EL, 2012).

Por fim, as perdas auditivas mistas são combinações de perdas auditivas condutivas e neurossensoriais. É o resultado de problemas na orelha externa, média e interna. As opções de tratamento variam desde medicamentos, cirurgia, aparelhos auditivos até implantes cocleares.

2.4 PRINCIPAIS MODELOS DE APARELHOS AUDITIVOS

Aparelhos auditivos são dispositivos utilizados para corrigir parcialmente as deficiências associadas a perdas auditivas (DILLON, 2001). A principal mudança que os aparelhos vêm sofrendo com o passar do tempo está relacionada com a redução de tamanho. A história desses dispositivos pode ser separada em cinco momentos: acústico, carbono, vácuo, transistor e digital. A última dessas eras é exatamente a que estamos presenciando hoje, prometendo avanços jamais esperados em eras anteriores.

Os principais elementos que compõem um aparelho auditivo digital são:

- Um ou mais microfones para converter o som em eletricidade;
- Um amplificador para aumentar a potência do sinal elétrico, de acordo com a perda do usuário;
- Um alto-falante miniatura chamado *receiver*, que converte de volta os sinais elétricos para sinais sonoros;

- Uma forma de acoplamento do som amplificado com o canal auditivo;
- Uma bateria para alimentar o sistema.

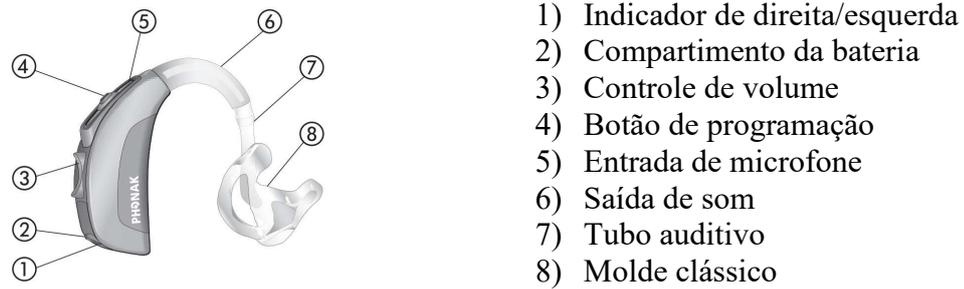
Existem basicamente dois grupos principais de aparelhos auditivos, os BTEs (*behind-the-ear*) e os ITE (*in-the-ear*), que diferem devido à forma com que são posicionados na orelha dos usuários. Cada um destes grupos apresenta ainda outras subdivisões que serão apresentadas a seguir.

2.4.1 Aparelhos BTE (*Behind-The-Ear*)

Os aparelhos chamados BTE (*behind-the-ear*), como o próprio nome sugere, utilizam a orelha para sustentação do aparelho. São muito robustos pois podem ser utilizados para todos os tipos de perdas auditivas, desde as suaves até as mais severas. O som gerado pelo aparelho é transmitido acusticamente via tubo até um molde personalizado conectado ao canal auditivo do indivíduo. A passagem das ondas sonoras pelo tubo até chegar à membrana timpânica pode afetar a qualidade do som. Existem diferentes modelos de molde que podem ser utilizados nesses aparelhos, dependendo da necessidade do paciente (BORISAGAR, THANKI e SEDANI, 2019).

Na Figura 14 é apresentada uma ilustração de um aparelho auditivo do tipo BTE modelo *Dalia* da fabricante *Phonak*[®], assim como a localização (esquerda) e indicação (direita) de seus principais componentes. Ainda na Figura 14 é possível observar que existe nesses aparelhos uma indicação de lado direito e lado esquerdo, evitando que usuários que utilizam aparelhos em ambas as orelhas invertam as posições. O botão de programação serve para modificar alguns parâmetros de configuração dos aparelhos de acordo com determinadas cenas acústicas visando melhorar a qualidade do som recebido pelo usuário. Pode ser utilizado um ou mais microfones, dependendo do modelo, ressaltando que aparelhos com mais microfones geralmente apresentam uma qualidade superior aos de microfone único. O som do aparelho é produzido pelo alto-falante, que nesse modelo encontra-se dentro do corpo do aparelho. A onda de saída passa pelo gancho, tubo auditivo e molde até chegar ao canal auditivo do usuário.

Figura 14 – Ilustração de um aparelho BTE modelo *Dalia* da fabricante *Phonak*.

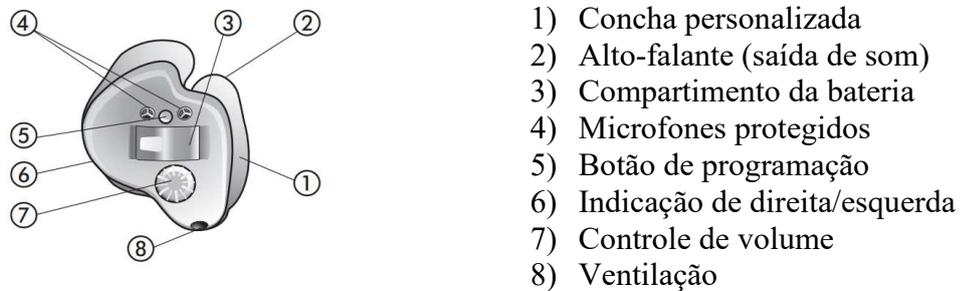


Fonte: (PHONAK, 2012).

2.4.2 Aparelhos ITE (*In-The-Ear*)

O próximo grupo de aparelhos auditivos mostrado é o ITE (*in-the-ear*), conhecido também como aparelho auditivo customizado. Esses aparelhos são fabricados individualmente para encaixar perfeitamente na orelha dos indivíduos e são indicados para os mais variados tipos de perdas auditivas (POPELKA *et al.*, 2016). Na Figura 15 é apresentada uma ilustração de um aparelho auditivo ITE clássico (esquerda), assim como uma descrição de cada um de seus componentes (direita).

Figura 15 – Ilustração de um aparelho ITE.



Fonte: (PHONAK, 2013).

2.4.3 Aparelhos Binaurais

Em mais de 80% de todos os pacientes com problemas auditivos, ambas as orelhas estão afetadas pela redução na habilidade auditiva. Isto significa que nesses casos os pacientes deveriam utilizar aparelhos auditivos em ambas as orelhas (WEIGMANN, 2014). Geralmente esses aparelhos são do tipo bilateral, ou seja, operam de maneira independente sem compartilhar

qualquer tipo de informação entre si. Nesse caso, torna-se muito difícil controlar ou preservar as pistas binaurais das fontes sonoras, afetando a localização das mesmas.

O termo aparelho binaural não se refere a um tipo específico de aparelho, mas sim à capacidade que um aparelho possui de se comunicar com o outro (contralateral). Essa funcionalidade permite que os procedimentos de redução de ruído utilizem informação proveniente de ambos os aparelhos cooperativamente. Isso aumenta a amostragem espacial dos sinais assim como o potencial dos métodos de redução ruído e a possibilidade de controle sobre a preservação da localização das fontes sonoras (DILLON, 2012).

2.5 MÉTRICAS DE AVALIAÇÃO DA FALA PROCESSADA

Nessa seção serão apresentadas as métricas objetivas utilizadas no trabalho para avaliar a qualidade e inteligibilidade da fala contaminada e processada pelos métodos de redução de ruído analisados.

O desempenho de um método de redução de ruído pode ser avaliado tanto com experimentos psicoacústicos de audição quanto com medidas objetivas. Experimentos psicoacústicos envolvem a comparação da fala original e processada por um grupo de pessoas que escutam e avaliam, pontuando dentro de uma escala predeterminada. A avaliação objetiva, por sua vez, é feita através de uma comparação matemática entre os sinais de fala original e processado (LOIZOU, 2013).

Qualidade é um atributo altamente subjetivo e por isso difícil de avaliar de maneira confiável. A qualidade é uma medida de *como* uma pessoa entende uma frase, que inclui atributos como “naturalidade”, “rugosidade”, entre outros. A qualidade possui muitas dimensões para se enumerar. A inteligibilidade, por sua vez, mede *o que* uma pessoa está entendendo, o que está ligado ao conteúdo das palavras proferidas, ou seja, à informação desejada.

As técnicas de processamento de sinais desenvolvidas nesse trabalho são aplicadas no domínio da frequência. A conversão desses sinais do domínio do tempo para o domínio da frequência é realizada através da transformada de *Fourier*. Devido ao comportamento não estacionário dos sinais de fala, a transformada de *Fourier* de tempo discreto (DTFT – do inglês *Discrete-Time Fourier Transform*) não é indicada, pois demanda longas janelas de análise que acabam omitindo importantes transições no conteúdo espectral dos sinais. Por esse motivo a transformada de *Fourier* de Tempo Curto (STFT – do inglês *Short-Time Fourier Transform*) é

utilizada, pois aplica a DTFT em curtos períodos de tempo (trechos) nos quais os sinais de fala podem ser considerados estacionários. A conversão de um sinal no domínio do tempo para o domínio da frequência utilizando a STFT é definida como:

$$y(\lambda, k) = \sum_{m=-\infty}^{\infty} y(m)w(\lambda - m)e^{-j\frac{2\pi}{N}km} \quad (3)$$

onde k é o índice que indica o bin, λ indica o trecho analisado, $y(m)$ é o sinal de entrada no domínio do tempo, $w(m)$ é a janela de análise, m o índice das amostras no tempo e N o total de bins da STFT.

2.5.1 Avaliação objetiva de qualidade

A década de 1990 foi marcada pelo crescimento do tráfego de sinais de fala em redes de pacotes como, por exemplo, a tecnologia de voz sobre IP (VOIP). Entretanto, nessa época não haviam medidas objetivas eficientes de qualidade da fala capazes de avaliar as distorções geradas por esse tipo de sistema de comunicação. Problemas como atraso variável da rede, perda de pacotes e as distorções geradas pelos codificadores-decodificadores (CODECs) geravam medidas de qualidade imprecisas para os algoritmos disponíveis na época.

Foi então que no ano 2000 a União Internacional de Telecomunicações (ITU – do inglês *International Telecommunication Union*) selecionou uma nova medida objetiva de qualidade mais confiável, levando em consideração vários problemas clássicos das redes telefônicas chaveadas por pacote. Assim nasceu a Avaliação Perceptual da Qualidade da Fala (PESQ – do inglês *Perceptual Evaluation of Speech Quality*) descrita em RIX *et al.*, (2001), e atualmente definida na recomendação P.862 da ITU-T (ITU, 2000). Como mencionado anteriormente, essa métrica foi concebida para avaliar a qualidade de sinais de fala da rede telefônica (*narrow band* - até 4 kHz). O resultado nessa versão de PESQ (PESQ *raw*) é apresentado em uma escala entre -0,5 e 4,5.

Alguns anos depois uma versão estendida de PESQ, visando avaliar sinais de banda larga, foi desenvolvida e regulamentada em (ITU, 2007). Nessa versão foi incluída uma função de mapeamento que permite uma comparação linear com uma famosa métrica subjetiva de qualidade chamada Pontuação Média de Opinião (MOS – do inglês *Mean Opinion Score*) (ITU, 2003) conhecida como MOS-LQO. O MOS utiliza uma escala de cinco níveis conhecida com

avaliação de categoria absoluta (ACR – do inglês *Absolute Category Rating*). A Tabela 3 apresenta a relação entre os resultados MOS-LQO e a escala ACR.

Tabela 3 – Mapeamento MOS-LQO na escala ACR.

MOS-LQO	Qualidade (ACR)
5	Excelente
4	Bom
3	Razoável
2	Ruim
1	Péssimo

2.5.2 Avaliação objetiva de inteligibilidade

No presente trabalho, a avaliação objetiva da inteligibilidade é realizada através da diferença entre a razão sinal-ruído ponderada pela inteligibilidade (ISNR – do inglês *Intelligibility Weighted SNR*) de entrada e da saída, proposta por GREENBERG, PETERSON e ZUREK (1993) e definida como:

$$\Delta ISNR = \sum_{i=1} I_i (SNR_{out}(i) - SNR_{in}(i)) \quad (4)$$

em que o índice i indexa cada uma das faixas de 1/3 de oitava de banda na qual a SNR será avaliada; $SNR_{in}(i)$ é a SNR do sinal de entrada do sistema de redução de ruído na faixa i ; e SNR_{out} é a SNR do sinal de saída na mesma faixa i . A função de importância da banda (I_i), que multiplica as diferenças de SNR, é definida em ACOUSTICAL SOCIETY OF AMERICA (1997). A ISNR reflete quanto a inteligibilidade da fala é melhorada pelos métodos de redução de ruído, sem levar em conta a distorção da fala (BENESTY, MAKINO e CHEN, 2005).

2.6 MÉTRICAS DE AVALIAÇÃO DA PRESERVAÇÃO DA LOCALIZAÇÃO

A seguir são apresentadas as métricas para a avaliação da preservação da localização espacial. As métricas selecionadas são utilizadas para medir os erros entre pistas binaurais dos sinais de entrada e saída do sistema de redução de ruído.

2.6.1 Erro de ITD

O erro de ITD ou ΔITD é calculado como a diferença entre a ITD dos sinais que chegam aos microfones de referência e a ITD dos sinais que são enviados aos alto-falantes dos aparelhos auditivos. O erro de ITD é calculado no domínio da frequência para todas as frequências até o limite (K) equivalente a 1,5 kHz.

$$\Delta ITD = \frac{1}{L} \sum_{\lambda=1}^L \sum_{k=1}^K \frac{|ITD_{out}(k, \lambda) - ITD_{in}(k, \lambda)|}{\pi} \quad (5)$$

em que $ITD_{in}(k, \lambda)$ e $ITD_{out}(k, \lambda)$ são, respectivamente, as ITDs dos sinais de entrada e saída no bin k e frame λ . O erro de ITD pode ser calculado tanto para a fala quanto para o ruído.

2.6.2 Erro de ILD

O erro de ILD é uma métrica calculada utilizando frequências acima de 1,5 kHz, de acordo com a teoria duplex. Quando seu valor é nulo, assume-se que ambos os sinais de entrada e saída de um aparelho auditivo possuem a mesma ILD. O erro de ILD é calculado como:

$$\Delta ILD = \frac{1}{L} \sum_{\lambda=1}^L \sum_{k=K+1}^N |10 \log_{10} ILD_{out}(k, \lambda) - 10 \log_{10} ILD_{in}(k, \lambda)| \quad (6)$$

em que $ILD_{in}(k, \lambda)$ e $ILD_{out}(k, \lambda)$ são, respectivamente, as ILDs dos sinais de entrada e saída no bin k e frame λ ; K é o índice associado à frequência de 1,5 kHz; e N é o número máximo de bins da representação no domínio da frequência. O erro de ILD pode ser calculado tanto para a fala quanto para o ruído.

2.6.3 Avaliação psicoacústica

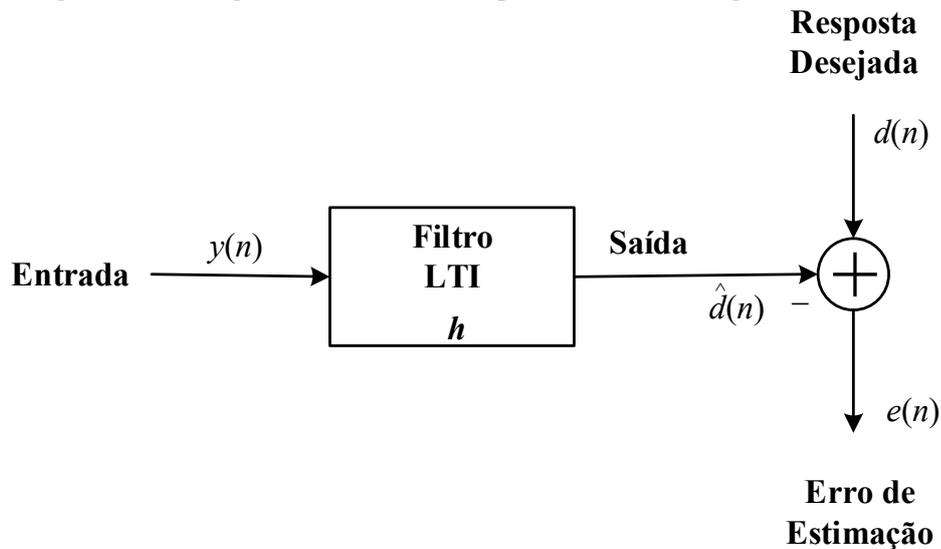
O procedimento de avaliação psicoacústica aplicado neste trabalho consiste em submeter um grupo de pessoas pré-selecionadas (voluntários) à audição, via fones de ouvido, de determinados arquivos de áudio contendo apenas ruído, processado ou não processado, e registrar resultados referentes à percepção do azimuth da fonte sonora. Para facilitar o

procedimento, uma ferramenta foi desenvolvida no programa MatLab[®] com uma interface gráfica intuitiva, submetendo o voluntário às etapas de aprendizado, treinamento e avaliação. O procedimento de seleção dos voluntários, assim como mais detalhes sobre a metodologia e o programa desenvolvido podem ser encontrados no Capítulo 4 e Apêndice A.

3 FILTRAGEM WIENER MULTICANAL PARA REDUÇÃO DE RUÍDO

Técnicas de redução de ruído são muito conhecidas na literatura e são exploradas nos mais variados cenários. Filtros seletores de frequência são um exemplo clássico, geralmente aplicados quando o sinal de interesse e o ruído interferente não ocupam a mesma banda de frequência. Entretanto, existem casos onde o sinal de interesse e o ruído estão com seus respectivos espectros sobrepostos. Nesses casos, técnicas de filtragem estatística são muito utilizadas e tentam separar informação e ruído através de suas características estatísticas para então reduzir/eliminar o ruído presente nos sinais de interesse. Uma técnica de filtragem estatística muito utilizada é chamada de filtragem Wiener (WIENER, 1949), possuindo uma ampla gama de aplicações que engloba além de redução de ruído, predição de sinais, cancelamento de eco e interpolação. O diagrama da Figura 16, exemplifica o problema de filtragem estatística.

Figura 16 – Diagrama em blocos do problema de filtragem estatística.



Como podemos observar na Figura 16, o sinal de entrada $y(n)$ é processado por um sistema linear e invariante no tempo (LTI – do inglês *Linear and Time Invariant*) h para produzir uma saída $\hat{d}(n)$ mais próxima possível (em algum sentido) do sinal desejado $d(n)$. Para tal, é necessário encontrar o conjunto de coeficientes ótimos do sistema que minimize o sinal de erro de estimação $e(n)$. O filtro que tem como objetivo minimizar o erro quadrático médio é conhecido como filtro de Wiener.

Assumindo-se sinais reais, a saída do filtro é dada pela convolução entre o sinal de entrada $y(n)$ e a resposta ao impulso do sistema LTI, resultando em

$$\hat{d}(n) = \sum_{k=0}^{N-1} h_k y(n-k) \quad (7)$$

ou na forma vetorial em

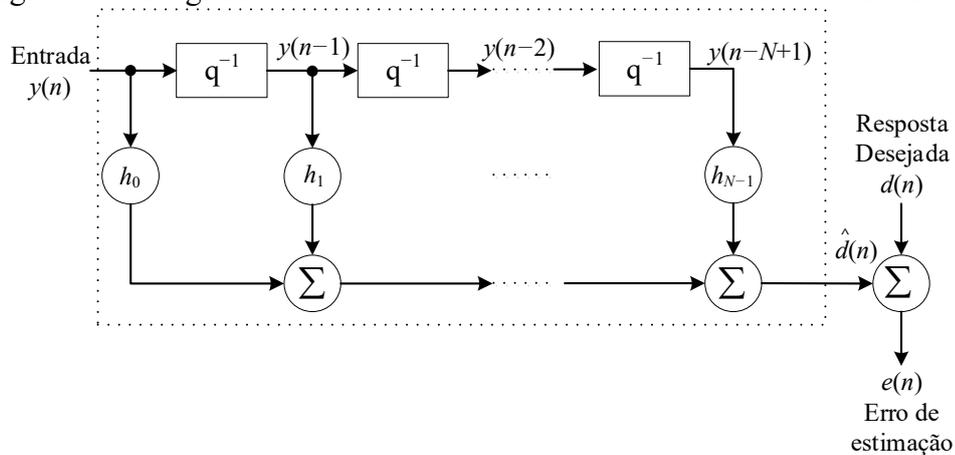
$$\hat{d}(n) = \mathbf{h}^T \mathbf{y}(n) \quad (8)$$

em que $\mathbf{h} = [h_0, h_1, \dots, h_{N-1}]^T$ representa o vetor de coeficientes do filtro de Wiener e $\mathbf{y}(n) = [y(n), y(n-1), \dots, y(n-N+1)]^T$ é o vetor das amostras de entrada no domínio do tempo. O erro de estimação é dado por

$$\begin{aligned} e(n) &= d(n) - \hat{d}(n) \\ &= d(n) - \mathbf{h}^T \mathbf{y}(n). \end{aligned} \quad (9)$$

Embora esses filtros possam ser implementados com respostas infinitas ao impulso (IIR – do inglês *Infinite Impulse Response*), estruturas com respostas finitas (FIR – do inglês *Finite Impulse Response*) são geralmente preferidas devido à sua inerente estabilidade. Um exemplo de uma versão FIR de um filtro de Wiener é apresentado na Figura 17. Os blocos denominados q^{-1} denotam atraso de uma unidade de tempo.

Figura 17 – Diagrama em blocos de um filtro de Wiener com N coeficientes.



A função custo do filtro de Wiener (J_{WF}) é definida como o erro médio quadrático $\mathbb{E}\{e^2(n)\}$, onde o operador $\mathbb{E}\{\cdot\}$ corresponde ao valor esperado da variável entre chaves, resultando em

$$\begin{aligned} J_{WF} &= \mathbb{E}\{e^2(n)\} = \mathbb{E}\{[d(n) - \mathbf{h}^T \mathbf{y}(n)]^2\} \\ &= \mathbb{E}\{d^2(n)\} - 2\mathbf{h}^T \mathbb{E}\{\mathbf{y}(n)d(n)\} + \mathbf{h}^T \mathbb{E}\{\mathbf{y}(n)\mathbf{y}^T(n)\}\mathbf{h} \\ &= \mathbb{E}\{d^2(n)\} - 2\mathbf{h}^T \mathbf{r}_{yd} + \mathbf{h}^T \mathbf{R}_{yy} \mathbf{h} \end{aligned} \quad (10)$$

em que o vetor $\mathbf{r}_{yd} = \mathbb{E}\{\mathbf{y}(n)d(n)\}$, de dimensão $(N \times 1)$, é o vetor de correlação cruzada entre o vetor de entrada e o sinal desejado e $\mathbf{R}_{yy} = \mathbb{E}\{\mathbf{y}(n)\mathbf{y}^T(n)\}$ é a matriz de autocorrelação $(N \times N)$ do sinal de entrada (si). A função custo J_{WF} terá seu valor mínimo quando todos os valores do vetor gradiente forem iguais a zero, de forma que

$$\frac{\partial J_{WF}}{\partial h_i} = 0 = 2\mathbb{E}\left\{e(n) \frac{\partial e(n)}{\partial h_i}\right\} \quad \text{para } i = 0, 1, \dots, N-1 \quad (11)$$

Utilizando-se (9) em (11) chega-se em

$$\frac{\partial J_{WF}}{\partial h_i} = -2\mathbb{E}\{e(n)y(n-i)\} = 0 \quad \text{para } i = 0, 1, \dots, N-1 \quad (12)$$

A equação (12), fornece uma visão das condições necessárias para que a função custo J_{WF} atinja seu valor mínimo. Em resumo, o erro de estimação $e(n)$ deve ser ortogonal ao sinal de entrada $y(n)$ para que o valor esperado do produto entre eles seja igual a zero. Este princípio é conhecido como princípio da ortogonalidade da filtragem linear ótima (BERTEIN e CESCHI, 2007). Por fim, arranjando-se (12) na forma vetorial encontra-se

$$\frac{\partial J_{WF}}{\partial \mathbf{h}} = -2\mathbf{r}_{yd} + 2\mathbf{h}^T \mathbf{R}_{yy} = 0 \quad (13)$$

Rearranjando-se (13), os coeficientes ótimos são dados por

$$\mathbf{h}_{opt} = \mathbf{R}_{yy}^{-1} \mathbf{r}_{yd} \quad (14)$$

A mesma ideia pode ser aplicada no domínio da frequência. Pode-se utilizar o mesmo diagrama apresentado na Figura 16. Entretanto, agora as variáveis são função não mais do tempo discreto, mas associadas aos índices de cada trecho (*frame*) e bin, assumindo-se que a transformação dos sinais é obtida através da STFT. Dessa forma, a saída $\hat{d}(\lambda, k)$ do *frame* λ -ésimo e k -ésimo bin é calculada através da multiplicação da STFT do sinal de entrada $y(\lambda, k)$ pela resposta em frequência do filtro $h(k)$ de forma que

$$\hat{d}(\lambda, k) = y(\lambda, k)h(k). \quad (15)$$

O erro de estimação, por sua vez, é calculado como,

$$\begin{aligned} e(\lambda, k) &= d(\lambda, k) - \hat{d}(\lambda, k) \\ &= d(\lambda, k) - y(\lambda, k)h(k) \end{aligned} \quad (16)$$

e a função custo resulta em

$$\begin{aligned} J_{WF}(k) &= \mathbb{E}\{e^2(\lambda, k)\} \\ &= \mathbb{E}\{[d(\lambda, k) - h(k)y(\lambda, k)]^2\} \\ &= \mathbb{E}\{[d(\lambda, k) - h(k)y(\lambda, k)]^* [d(\lambda, k) - h(k)y(\lambda, k)]\} \\ &= \mathbb{E}\{|d(\lambda, k)|^2\} - h(k)\mathbb{E}\{d^*(\lambda, k)y(\lambda, k)\} \\ &\quad - h^*(k)\mathbb{E}\{y^*(\lambda, k)d(\lambda, k)\} \\ &\quad + |h(k)|^2 \mathbb{E}\{|y(\lambda, k)|^2\} \end{aligned} \quad (17)$$

em que a densidade espectral de potência do sinal de entrada $y(n)$ e do sinal desejado $d(n)$ são dadas respectivamente por

$$P_y(k) = \mathbb{E}\{|y(\lambda, k)|^2\}, \quad P_d(k) = \mathbb{E}\{|d(\lambda, k)|^2\} \quad (18)$$

e a densidade espectral de potência cruzada entre os sinais $y(n)$ e $d(n)$ é

$$P_{yd}(k) = P_{dy}^*(k) = \mathbb{E}\{y(\lambda, k)d^*(\lambda, k)\}. \quad (19)$$

O índice λ foi suprimido em função da suposição de estacionariedade dos sinais. Reescrevendo (17), chega-se na função custo,

$$\begin{aligned} J_{\text{WF}}(k) &= \mathbb{E}\{|e(k)|^2\} \\ &= P_d(k) - h(k)P_{yd}(k) - h^*(k)P_{dy}(k) + |h(k)|^2 P_y(k). \end{aligned} \quad (20)$$

Derivando-se (20) e igualando a zero, chega-se enfim na equação geral do filtro de Wiener no domínio da frequência

$$h(k) = \frac{P_{dy}(k)}{P_y(k)}. \quad (21)$$

É importante destacar que os coeficientes do filtro $h(k)$ são geralmente complexos devido aos valores de $P_{dy}(k)$ serem também geralmente complexos. O processo de filtragem utilizado nesse trabalho é realizado no domínio da frequência. Dessa forma, a partir da próxima seção, a menos que estritamente estabelecido, todas as variáveis apresentadas referem-se à sua representação no domínio da frequência. A nomenclatura utilizada estabelece que letras maiúsculas e minúsculas em negrito representam matrizes e vetores, respectivamente, enquanto que letras minúsculas em itálico são associadas a variáveis escalares e maiúsculas em itálico são constantes.

3.1 FILTRO DE WIENER MONOCANAL PARA REDUÇÃO DE RUÍDO

Como citado no início do capítulo, a técnica de filtragem Wiener possui aplicações em problemas de diferentes naturezas. Nesta seção o foco estará na redução de ruído, uma vez que esta é a aplicação alvo do presente trabalho. Para a particularização da estrutura genérica da filtragem Wiener apresentada na Figura 17 para a aplicação em redução de ruído, duas considerações devem ser feitas. A primeira delas é

$$y(\lambda, k) = x(\lambda, k) + v(\lambda, k) \quad (22)$$

que significa que o sinal de entrada no domínio da frequência $y(\lambda, k)$ é composto por uma parcela chamada de sinal de interesse (fala), dada por $x(\lambda, k)$, e um ruído aditivo, chamado de $v(\lambda, k)$. A segunda consideração é com relação ao sinal desejado $d(\lambda, k)$, mostrado na Figura 16, onde

$$d(\lambda, k) = x(\lambda, k) \quad (23)$$

ou seja, o sinal desejado é igual ao sinal de fala.

Assumindo que a fala e o ruído são variáveis aleatórias independentes de média zero e substituindo (22) e (23) em (19), chega-se na definição da densidade espectral cruzada, entre o sinal de entrada e o sinal desejado, dada por

$$\begin{aligned} P_{dy}(k) &= \mathbb{E}\{x(\lambda, k)[x(\lambda, k) + v(\lambda, k)]^*\} \\ &= \mathbb{E}\{x(\lambda, k)x^*(\lambda, k)\} + \mathbb{E}\{x(\lambda, k)v^*(\lambda, k)\} \\ &= P_x(k) \end{aligned} \quad (24)$$

que é igual à densidade espectral de potência do sinal de fala. Substituindo agora (22) e (23) em (18), têm-se

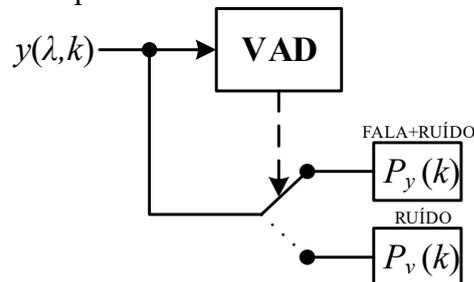
$$\begin{aligned} P_y(k) &= \mathbb{E}\{[x(\lambda, k) + v(\lambda, k)][x(\lambda, k) + v(\lambda, k)]^*\} \\ &= \mathbb{E}\{x(\lambda, k)x^*(\lambda, k)\} + \mathbb{E}\{v(\lambda, k)v^*(\lambda, k)\} \\ &\quad + \mathbb{E}\{x(\lambda, k)v^*(\lambda, k)\} + \mathbb{E}\{x(\lambda, k)v^*(\lambda, k)\} \\ &= P_x(k) + P_v(k) \end{aligned} \quad (25)$$

e, portanto, a densidade espectral de potência de entrada é igual à soma das densidades espectrais de potência do sinal de fala e do ruído. Finalmente, substituindo as equações (24) e (25) em (21), obtém-se a equação do filtro de Wiener para redução de ruído:

$$h(k) = \frac{P_x(k)}{P_x(k) + P_v(k)} \quad (26)$$

Em aplicações práticas, não é possível acessar diretamente o sinal de fala limpo $x(\lambda, k)$ e, por consequência, não é possível calcular diretamente a densidade espectral de potência do sinal de fala $P_x(k)$ exigida para calcular os coeficientes do filtro, conforme (26). Todavia, existem outras maneiras de calcular esse valor. Em aplicações com sinais de fala, com uma fonte de ruído ininterrupta, por exemplo, existem basicamente dois momentos, um em que há apenas o ruído ($y(\lambda, k) = v(\lambda, k)$) e outro em que há ruído e fala ($y(\lambda, k) = x(\lambda, k) + v(\lambda, k)$). Sendo assim, uma prática comum para a estimação das densidades espectrais de potência desejadas baseada na utilização de um detector de fala (VAD – do inglês *Voice Activity Detector*) que funciona como um classificador ora indicando a presença de apenas ruído ou fala e ruído. Nos *frames* em que haja apenas ruído, apenas a densidade espectral de potência do ruído $P_v(k)$ é estimada, quando houver fala e ruído apenas a densidade espectral do sinal de entrada $P_y(k)$ é estimada.

Figura 18 – Utilização de um detector de fala para selecionar a densidade espectral de potência a ser estimada.



Com as densidades espectrais de potência do ruído e do sinal de entrada estimadas, assumindo que fala e ruído são independentes, é possível estimar a densidade espectral de potência da fala a partir de

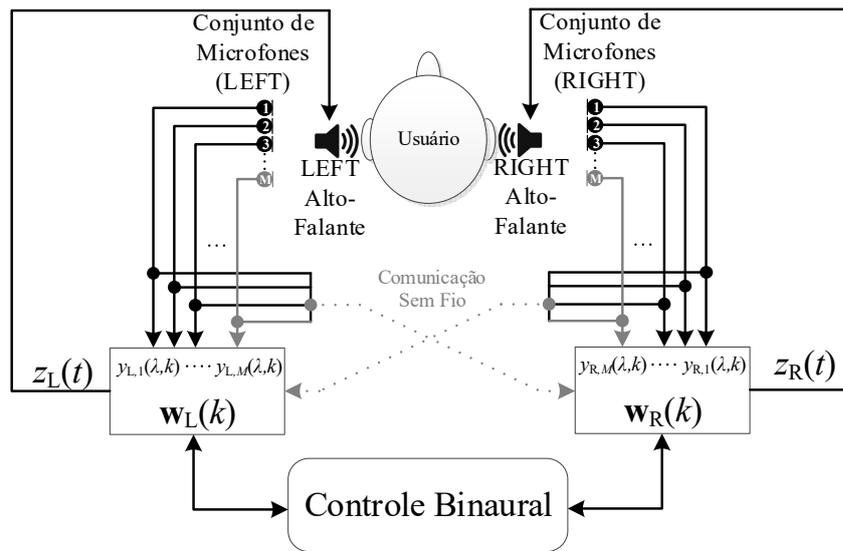
$$P_x(k) = P_y(k) - P_v(k). \quad (27)$$

3.2 FILTRO DE WIENER MULTICANAL PARA REDUÇÃO DE RUÍDO

Nas seções anteriores foram apresentadas as derivações da filtragem Wiener com uma única entrada (monocanal) nos domínios do tempo e frequência, assim como a particularização da estrutura visando a redução de ruído no domínio da frequência. O objetivo desta seção é

ampliar essas definições para sistemas mais complexos compostos por múltiplos canais de entrada (multicanal) aplicados especificamente em aparelhos auditivos binaurais. A Figura 19 apresenta em blocos a configuração básica de um sistema de redução de ruído em um aparelho auditivo binaural multicanal. Cada aparelho auditivo possui um conjunto de M microfones. Os sinais captados por ambos conjuntos de microfones são compartilhados pelos dois aparelhos auditivos. Essa troca de informação ocorre através de comunicação sem fio e tem seu atraso desconsiderado nesse trabalho. O bloco denominado “Controle Binaural” é responsável por processar todas as informações dos $2M$ microfones para gerar os coeficientes dos filtros do aparelho auditivo esquerdo $\mathbf{w}_L(k)$ e direito $\mathbf{w}_R(k)$. Por fim, os sinais filtrados $z_L(\lambda, k)$ e $z_R(\lambda, k)$ são enviados aos seus respectivos alto-falantes após a reconstrução do sinal no domínio do tempo utilizando, por exemplo, a técnica de *overlap-and-add* (CROCHIERE, 1980).

Figura 19 – Configuração de um aparelho auditivo binaural.



O sinal do n -ésimo microfone do aparelho auditivo esquerdo (L) no domínio da frequência pode ser expresso como

$$y_{L,m}(\lambda, k) = x_{L,m}(\lambda, k) + v_{L,m}(\lambda, k), \quad m = 1, \dots, M \quad (28)$$

assim como o sinal que chega aos microfones do aparelho auditivo direito (R) é definido como

$$y_{R,m}(\lambda, k) = x_{R,m}(\lambda, k) + v_{R,m}(\lambda, k), \quad m = 1, \dots, M \quad (29)$$

Por fim, os vetores \mathbf{y}_L e \mathbf{y}_R de dimensão M são empilhados para dar origem ao vetor de entrada binaural $\mathbf{y}(k)$ de dimensão $2M$

$$\mathbf{y}_L(\lambda, k) = \begin{bmatrix} y_{L,1}(\lambda, k) \\ y_{L,2}(\lambda, k) \\ \vdots \\ y_{L,M}(\lambda, k) \end{bmatrix}, \quad \mathbf{y}_R(\lambda, k) = \begin{bmatrix} y_{R,1}(\lambda, k) \\ y_{R,2}(\lambda, k) \\ \vdots \\ y_{R,M}(\lambda, k) \end{bmatrix}, \quad \mathbf{y}(\lambda, k) = \begin{bmatrix} \mathbf{y}_L(\lambda, k) \\ \mathbf{y}_R(\lambda, k) \end{bmatrix} \quad (30)$$

O vetor do sinal de entrada pode ser escrito como $\mathbf{y}(\lambda, k) = \mathbf{x}(\lambda, k) + \mathbf{v}(\lambda, k)$, em que $\mathbf{x}(\lambda, k)$ e $\mathbf{v}(\lambda, k)$ são definidos de maneira similar ao vetor $\mathbf{y}(\lambda, k)$. As matrizes de coerência do sinal de entrada $\Phi_{yy}(k)$ e do ruído $\Phi_{vv}(k)$ são estimadas com o auxílio do VAD e definidas como

$$\begin{aligned} \Phi_{yy}(k) &= \mathbb{E}\{\mathbf{y}(\lambda, k)\mathbf{y}^H(\lambda, k)\}, \\ \Phi_{vv}(k) &= \mathbb{E}\{\mathbf{v}(\lambda, k)\mathbf{v}^H(\lambda, k)\}. \end{aligned} \quad (31)$$

Considera-se que o sinal de fala é independente do sinal de ruído, o que permite que a matriz de coerência da fala seja estimada como

$$\Phi_{xx}(k) = \Phi_{yy}(k) - \Phi_{vv}(k) \quad (32)$$

Como mostrado na Figura 19 e em (30) os aparelhos auditivos binaurais são geralmente compostos por múltiplos microfones. Geralmente os microfones frontais ($m=1$) na Figura 19 são escolhidos como os microfones de referência em cada aparelho (esquerda/direita). Seus respectivos sinais são definidos como

$$y_{L,REF}(\lambda, k) = \mathbf{q}_L^T \mathbf{y}(\lambda, k), \quad y_{R,REF}(\lambda, k) = \mathbf{q}_R^T \mathbf{y}(\lambda, k) \quad (33)$$

em que \mathbf{q}_L e \mathbf{q}_R são vetores de seleção, de dimensão $2M$, dos sinais dos microfones de referência. Esses vetores são compostos por $2M-1$ elementos iguais a zero e apenas um elemento igual a 1. Por exemplo, se considerarmos $M = 3$, temos ambos os vetores como

$$\mathbf{q}_L^T = [1, 0, 0, 0, 0, 0], \quad \mathbf{q}_R^T = [0, 0, 0, 1, 0, 0]. \quad (34)$$

Os sinais de saída $z_L(\lambda, k)$ e $z_R(\lambda, k)$ nos alto-falantes esquerdo e direito, respectivamente, são gerados através da filtragem e soma de todos os sinais dos microfones de ambos aparelhos auditivos,

$$z_L(\lambda, k) = \mathbf{w}_L^H(k) \mathbf{y}(\lambda, k), \quad z_R(\lambda, k) = \mathbf{w}_R^H(k) \mathbf{y}(\lambda, k) \quad (35)$$

em que $\mathbf{w}_L(k)$ e $\mathbf{w}_R(k)$ são os vetores dos coeficientes complexos de dimensão $2M$.

O filtro de Wiener multicanal (MWF) produz uma estimativa de mínimo erro quadrático médio do componente de fala no microfone de referência de cada aparelho auditivo ($\hat{x}_{L,REF}(\lambda, k)$ e $\hat{x}_{R,REF}(\lambda, k)$), por isso simultaneamente reduz ruído e limita a distorção da fala. A função custo do filtro de Wiener multicanal (MWF) é

$$J_{MWF}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = \mathbb{E} \left\{ \left\| \begin{array}{l} x_{L,REF}(\lambda, k) - \mathbf{w}_L^H(k) \mathbf{y}(\lambda, k) \\ x_{R,REF}(\lambda, k) - \mathbf{w}_R^H(k) \mathbf{y}(\lambda, k) \end{array} \right\|^2 \right\} \quad (36)$$

em que $\|\cdot\|^2$ é a norma Euclideana. Manipulando-se (36) chega-se em

$$\begin{aligned} J_{MWF}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= \phi_{x_L}(k) - \mathbf{w}_L^H(k) \boldsymbol{\phi}_{x_L}(k) - \boldsymbol{\phi}_{x_L}^H(k) \mathbf{w}_L(k) \\ &\quad + \mathbf{w}_L^H(k) \boldsymbol{\Phi}_{yy}(k) \mathbf{w}_L(k) + \phi_{x_R}(k) \\ &\quad - \mathbf{w}_R^H(k) \boldsymbol{\phi}_{x_R}(k) - \boldsymbol{\phi}_{x_R}^H(k) \mathbf{w}_R(k) \\ &\quad + \mathbf{w}_R^H(k) \boldsymbol{\Phi}_{yy}(k) \mathbf{w}_R(k) \end{aligned} \quad (37)$$

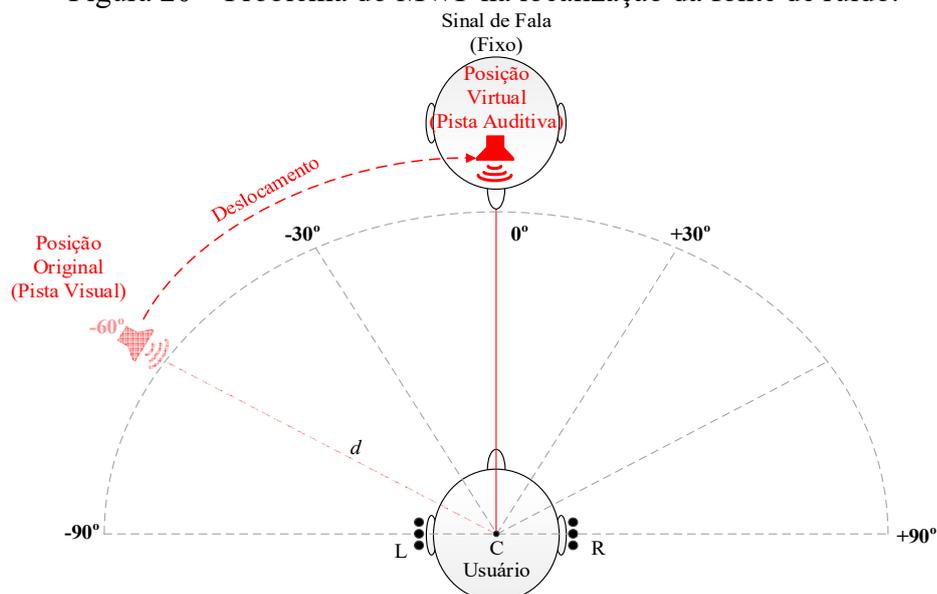
em que $\phi_{x_l}(k) = \mathbb{E} \{|x_{l,REF}(\lambda, k)|^2\} = \mathbf{q}_l^T \boldsymbol{\Phi}_{xx}(k) \mathbf{q}_l$, para $l \in \{L, R\}$, são as potências, no bin k , do sinal de entrada nos microfones frontais esquerdo e direito, respectivamente, e $\boldsymbol{\phi}_{x_l}(k) = \mathbb{E} \{\mathbf{x}(\lambda, k) x_{l,REF}^*(\lambda, k)\} = \boldsymbol{\Phi}_{xx}(k) \mathbf{q}_l$.

3.3 EFEITOS DA FILTRAGEM WIENER MULTICANAL NAS PISTAS BINAURAIS DA FALA E RUÍDO

Como mostrado na seção anterior, o filtro de Wiener multicanal estima o sinal de fala no microfone de referência de cada aparelho auditivo reduzindo assim o ruído e limitando a distorção da fala. Em outras palavras, o MWF naturalmente preserva as pistas binaurais do sinal de fala. Entretanto, em DOCCLO, KLASSEN, *et al.* (2006) foi mostrado que as pistas binaurais (ITD e ILD) do ruído processado pelo MWF, no caso de um cenário acústico com duas fontes pontuais, sendo uma de fala e outra de ruído (e relação sinal ruído positiva), tendem às pistas binaurais da fonte de fala. Na prática isso significa dizer que o MWF “arrasta” a fonte de ruído para o azimute da fonte de fala, o que obviamente é um efeito indesejado e necessita ser corrigido. Tal efeito gera um descasamento entre as pistas visuais e auditivas da fonte de ruído percebida pelo usuário do aparelho auditivo que, além de desconfortável, pode ser até mesmo perigoso em situações cotidianas.

O diagrama da Figura 20 exemplifica o problema ocasionado pelo MWF com o ruído processado. Podemos observar que a cena acústica real é composta por uma fonte de fala com azimute em 0° e uma fonte de ruído no azimute original de -60° . Entretanto, o processamento efetuado pelo MWF modifica apenas as pistas binaurais da fonte de ruído deixando-as próximas das pistas binaurais da fonte de fala. Isto faz com que pareça que a fonte de ruído está localizada no mesmo azimute da fonte de fala (posição virtual), ou seja, com azimute de 0° .

Figura 20 – Problema do MWF na localização da fonte de ruído.



3.4 FILTRAGEM WIENER MULTICANAL COM TÉCNICAS DE PRESERVAÇÃO DAS PISTAS BINAURAS DA FONTE DE RUÍDO

A seção anterior descreveu o problema gerado pelo MWF que muda a sensação subjetiva da posição real da fonte de ruído. Esse efeito é obviamente indesejado e deve ser corrigido, ou seja, a sensação subjetiva da fonte de ruído deve ser restabelecida.

Para tanto, existem dois tipos de estratégias disponíveis na literatura que operam com o MWF no domínio da frequência. No primeiro tipo, a própria função custo J_{MWF} (equação (37)) é alterada. No segundo, é adicionada uma função custo auxiliar ($J_{AUXILIAR}$) à função custo original do J_{MWF} . Esse segundo grupo gera uma nova função custo total cujo termo adicional tem como objetivo controlar alterações na modificação das pistas binaurais da fonte de ruído, como mostra a equação (38).

$$J_{TOTAL}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = J_{MWF}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) + J_{AUXILIAR}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \quad (38)$$

Geralmente, essas funções custo auxiliares são multiplicadas por parâmetros que ponderam a importância de sua atuação dentro da função total. Isso permite estabelecer o quanto essa função custo adicional irá influenciar o sinal processado. Diversos resultados apresentados neste trabalho baseiam-se na variação desses parâmetros de controle. A seguir são detalhadas algumas das principais técnicas que visam a preservação do azimuth da fonte de ruído e que foram brevemente apresentadas no Capítulo 1.

3.4.1 MWF-SDW (*Speech Distortion Weighted*)

Como já destacado anteriormente, a aplicação do filtro de Wiener multicanal (MWF) não altera substancialmente as pistas binaurais da fala. Visando preservar também as pistas do ruído, foi proposta uma técnica em (SPRIET, MOONEN e WOUTERS, 2004) que fornece um compromisso entre distorção da fala e redução de ruído. O MWF com *Speech Distortion Weighted* (SDW) minimiza a soma ponderada da energia residual do ruído e a energia de distorção do sinal de fala. A função custo MWF-SDW é

$$\begin{aligned}
& J_{\text{MWF-SDW}}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \\
&= \mathbb{E} \left\{ \left\| \begin{array}{l} x_{\text{L,REF}}(\lambda, k) - \mathbf{w}_L^H(k) \mathbf{x}(\lambda, k) \\ x_{\text{R,REF}}(\lambda, k) - \mathbf{w}_R^H(k) \mathbf{x}(\lambda, k) \end{array} \right\|^2 + \mu \left\| \begin{array}{l} \mathbf{w}_L^H(k) \mathbf{v}(\lambda, k) \\ \mathbf{w}_R^H(k) \mathbf{v}(\lambda, k) \end{array} \right\|^2 \right\} \quad (39)
\end{aligned}$$

em que μ é o parâmetro que controla a relação entre a distorção da fala e a redução de ruído.

3.4.2 MWF- η (*Partial Estimation of Noise Component*)

Como mencionado anteriormente, essa extensão foi proposta em KLASSEN *et al.* (2005) para preservar parcialmente as pistas binaurais do ruído. O objetivo é produzir uma estimativa com mínimo erro quadrático médio (MMSE – do inglês *Minimum Mean Squared Error*) entre o sinal desejado (igual à soma da componente de fala com uma versão ponderada da componente de ruído) e a saída do filtro. A função custo é definida como

$$\begin{aligned}
& J_{\text{MWF-}\mu}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \\
&= \mathbb{E} \left\{ \left\| \begin{array}{l} [x_{\text{L,REF}}(\lambda, k) + \lambda v_{\text{L,REF}}(\lambda, k)] - \mathbf{w}_L^H(k) \mathbf{y}(\lambda, k) \\ [x_{\text{R,REF}}(\lambda, k) + \lambda v_{\text{R,REF}}(\lambda, k)] - \mathbf{w}_R^H(k) \mathbf{y}(\lambda, k) \end{array} \right\|^2 \right\} \quad (40)
\end{aligned}$$

para $0 \leq \lambda \leq 1$. Quando $\lambda = 0$, a função custo se torna igual à função custo $J_{\text{MWF}}(k, \mathbf{w}_L(k), \mathbf{w}_R(k))$. Entretanto, quando $\lambda = 1$, o filtro ótimo é um vetor de zeros com apenas um elemento igual a 1, que significa que não haverá redução de ruído, mas sim a manutenção do sinal de fala contaminada.

3.4.3 MWF-ILD (*Interaural Level Difference*)

O método MWF-ILD visa minimizar o valor esperado da norma quadrada da diferença entre a ILD de entrada e a ILD de saída do ruído. Essa abordagem faz parte do grupo que utiliza funções custo específicas adicionadas à função custo do MWF. A equação (41) apresenta a função custo total.

$$J_{\text{TOTAL}}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = J_{\text{MWF}}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) + \alpha J_{\text{ILD}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \quad (41)$$

O parâmetro α é utilizado para controlar a importância da função custo J_{ILD} dentro da função custo total. A equação (42) define a função custo destinada à preservação da ILD do ruído.

$$\begin{aligned} J_{ILD}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= \mathbb{E} \left\{ \left| \Delta \text{ILD}^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \right|^2 \right\} \\ &= \mathbb{E} \left\{ \left| \text{ILD}_{\text{out}}^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) - \text{ILD}_{\text{in}}^v(\lambda, k) \right|^2 \right\} \end{aligned} \quad (42)$$

A diferença de nível interaural instantânea (em dB) do ruído na entrada, calculada a partir dos microfones de referências, é definida em RASPAUD, VISTE e EVANGELISTA (2010) como

$$\begin{aligned} \text{ILD}_{\text{in}}^v(\lambda, k) &= 20 \log_{10} \left| \frac{v_{L,\text{REF}}(\lambda, k)}{v_{R,\text{REF}}(\lambda, k)} \right| \\ &= 10 \log_{10} \left| \frac{v_{L,\text{REF}}(\lambda, k)}{v_{R,\text{REF}}(\lambda, k)} \right|^2 \end{aligned} \quad (43)$$

e a diferença de nível interaural instantânea (em dB) do ruído de saída é definida como

$$\begin{aligned} \text{ILD}_{\text{out}}^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= 20 \log_{10} \left| \frac{z_{L,\text{REF}}(\lambda, k)}{z_{R,\text{REF}}(\lambda, k)} \right| \\ &= 10 \log_{10} \left| \frac{z_{L,\text{REF}}(\lambda, k)}{z_{R,\text{REF}}(\lambda, k)} \right|^2. \end{aligned} \quad (44)$$

A partir das definições apresentadas em (43) e (44) chega-se na definição da variação da ILD instantânea do ruído (em dB).

$$\Delta \text{ILD}^v(\lambda, k) = 10 \log_{10} \left| \frac{z_{L,\text{REF}}(\lambda, k)}{z_{R,\text{REF}}(\lambda, k)} \right|^2 - 10 \log_{10} \left| \frac{v_{L,\text{REF}}(\lambda, k)}{v_{R,\text{REF}}(\lambda, k)} \right|^2 \quad (45)$$

Para a definição da função custo da ILD, algumas aproximações são necessárias. Em DOCLO *et al.* (2005a) a primeira aproximação apresentada foi a substituição da operação do logaritmo, mostrada em (45) por uma aproximação de primeira ordem de sua série de Taylor ($\log_{10}(x) \approx (x-1)/\log_e(10)$). Com essa aproximação chega-se em:

$$J_{\text{ILD}}^v(\lambda, k) \cong \frac{100}{(\log_e 10)^2} \mathbb{E} \left[\left(\frac{|z_{\text{L,REF}}(\lambda, k)|^2}{|z_{\text{R,REF}}(\lambda, k)|^2} - 1 \right) - \left(\frac{|v_{\text{L,REF}}(\lambda, k)|^2}{|v_{\text{R,REF}}(\lambda, k)|^2} - 1 \right) \right]^2 \quad (46)$$

$$\cong \frac{100}{(\log_e 10)^2} \mathbb{E} \left| \frac{|z_{\text{L,REF}}(\lambda, k)|^2}{|z_{\text{R,REF}}(\lambda, k)|^2} - \frac{|v_{\text{L,REF}}(\lambda, k)|^2}{|v_{\text{R,REF}}(\lambda, k)|^2} \right|^2.$$

Em sequência, aproxima-se o valor esperado do módulo ao quadrado pelo módulo ao quadrado do valor esperado, de tal forma que:

$$J_{\text{ILD}}^v(\lambda, k) = \frac{100}{(\log_e 10)^2} \left| \mathbb{E} \left\{ \frac{|z_{\text{L,REF}}(\lambda, k)|^2}{|z_{\text{R,REF}}(\lambda, k)|^2} \right\} - \mathbb{E} \left\{ \frac{|v_{\text{L,REF}}(\lambda, k)|^2}{|v_{\text{R,REF}}(\lambda, k)|^2} \right\} \right|^2. \quad (47)$$

Além disso, os valores médios das razões são separados no valor esperado do numerador e denominador.

$$\mathbb{E} \left\{ \frac{|v_{\text{L,REF}}(\lambda, k)|^2}{|v_{\text{R,REF}}(\lambda, k)|^2} \right\} \cong \frac{\mathbb{E} \{ |v_{\text{L,REF}}(\lambda, k)|^2 \}}{\mathbb{E} \{ |v_{\text{R,REF}}(\lambda, k)|^2 \}} = \frac{\mathbf{q}_{\text{L}}^T \Phi_{\text{vv}}(k) \mathbf{q}_{\text{L}}}{\mathbf{q}_{\text{R}}^T \Phi_{\text{vv}}(k) \mathbf{q}_{\text{R}}} \quad (48)$$

e da mesma forma

$$\mathbb{E} \left\{ \frac{|z_{\text{L,REF}}(\lambda, k)|^2}{|z_{\text{R,REF}}(\lambda, k)|^2} \right\} \cong \frac{\mathbb{E} \{ |z_{\text{L,REF}}(\lambda, k)|^2 \}}{\mathbb{E} \{ |z_{\text{R,REF}}(\lambda, k)|^2 \}} = \frac{\mathbf{w}_{\text{L}}^H \Phi_{\text{vv}}(k) \mathbf{w}_{\text{L}}}{\mathbf{w}_{\text{R}}^H \Phi_{\text{vv}}(k) \mathbf{w}_{\text{R}}}. \quad (49)$$

Por fim, negligenciando-se o fator de escala, a função custo proposta em DOCLO *et al.* (2005a) foi apresentada como:

$$J_{\text{ILD}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = \left| \frac{\mathbf{w}_L^H \Phi_{\text{vv}}(k) \mathbf{w}_L - \mathbf{q}_L^T \Phi_{\text{vv}}(k) \mathbf{q}_L}{\mathbf{w}_R^H \Phi_{\text{vv}}(k) \mathbf{w}_R - \mathbf{q}_R^T \Phi_{\text{vv}}(k) \mathbf{q}_R} \right|^2. \quad (50)$$

Em (COSTA e NAYLOR, 2014) outra aproximação do logaritmo foi apresentada, obtida através da função tangente hiperbólica inversa. A função custo proposta nesse trabalho é dada por

$$J_{\text{ILD}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = \left| \frac{\mathbf{w}_L^H \Phi_{\text{vv}}(k) \mathbf{w}_L \mathbf{q}_R^T \Phi_{\text{vv}}(k) \mathbf{q}_R - \mathbf{w}_R^H \Phi_{\text{vv}}(k) \mathbf{w}_R \mathbf{q}_L^T \Phi_{\text{vv}}(k) \mathbf{q}_L}{\mathbf{w}_L^H \Phi_{\text{vv}}(k) \mathbf{w}_L \mathbf{q}_R^T \Phi_{\text{vv}}(k) \mathbf{q}_R + \mathbf{w}_R^H \Phi_{\text{vv}}(k) \mathbf{w}_R \mathbf{q}_L^T \Phi_{\text{vv}}(k) \mathbf{q}_L} \right|^2. \quad (51)$$

Como comentado em capítulos anteriores, segundo a teoria duplex, apenas frequências acima de 1,5 kHz possuem valores significativos de ILD. Por esse motivo, esta função custo (com as suas devidas modificações) é aplicada apenas em bins de frequência acima desse valor. Essa função custo também pode ser aplicada à fala, entretanto, como o MWF preserva naturalmente as pistas acústicas da fala, a mesma não será utilizada neste trabalho.

3.4.4 MWF-ITD (*Interaural Time Difference*)

O método MWF-ITD é da mesma família apresentada na seção anterior, também utilizada em conjunto com a função custo da MWF. A função custo total é determinada como

$$J_{\text{TOTAL}}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = J_{\text{MWF}}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) + \beta J_{\text{ITD}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \quad (52)$$

em que β é o parâmetro que controla o peso da função J_{ITD} dentro da função total.

O objetivo dessa abordagem é preservar diferença de tempo interaural (ITD) do ruído.

$$\begin{aligned} J_{\text{ITD}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= \mathbb{E} \left\{ \left| \Delta \text{ITD}_v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \right|^2 \right\} \\ &= \mathbb{E} \left\{ \left| \text{ITD}_{\text{out}}^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) - \text{ITD}_{\text{in}}^v(\lambda, k) \right|^2 \right\} \end{aligned} \quad (53)$$

A diferença de tempo interaural instantânea do ruído (em segundos), calculada a partir dos microfones de referência é definida em RASPAUD, VISTE e EVANGELISTA (2010) como

$$\text{ITD}_{\text{in}}^v(\lambda, k) = \frac{1}{\omega_k} \left(\angle \frac{v_{L,REF}(\lambda, k)}{v_{R,REF}(\lambda, k)} + 2\pi p(k) \right) \quad (54)$$

em que \angle significa fase. A diferença de tempo interaural instantânea do ruído (em segundos) nos alto-falantes é definida como

$$\text{ITD}_{\text{out}}^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = \frac{1}{\omega_k} \left(\angle \frac{z_{L,REF}(\lambda, k)}{z_{R,REF}(\lambda, k)} + 2\pi p(k) \right) \quad (55)$$

em que ω é a frequência angular do bin k e o valor inteiro p é o fator de *unwrapping* da fase conhecido *a priori*. Esse fator p é usado acima de certas frequências para evitar que a fase se torne ambígua e depende principalmente da distância entre as orelhas.

Em DOCLO *et al.* (2005b) o valor médio da ITD de entrada foi definido como

$$\begin{aligned} \text{ITD}_{\text{in}}^v(k) &= \frac{1}{\omega_k} \mathbb{E} \left\{ \angle \frac{v_{L,REF}(\lambda, k)}{v_{R,REF}(\lambda, k)} \cdot \frac{v_{R,REF}^*(\lambda, k)}{v_{R,REF}^*(\lambda, k)} \right\} \\ &= \frac{1}{\omega_k} \mathbb{E} \left\{ \angle \frac{v_{L,REF}(\lambda, k) \cdot v_{R,REF}^*(\lambda, k)}{|v_{R,REF}(\lambda, k)|^2} \right\} \\ &= \frac{1}{\omega_k} \mathbb{E} \left\{ \angle v_{L,REF}(\lambda, k) v_{R,REF}^*(\lambda, k) \right\} \\ &\cong \frac{1}{\omega_k} \angle \mathbb{E} \left\{ v_{L,REF}(\lambda, k) v_{R,REF}^*(\lambda, k) \right\} \\ &= \frac{1}{\omega_k} \angle \mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R \end{aligned} \quad (56)$$

e o valor médio da ITD de saída como

$$\begin{aligned}
ITD_{out}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= \frac{1}{\omega_k} \mathbb{E} \left\{ \angle \frac{z_{L,REF}(\lambda, k)}{z_{R,REF}(\lambda, k)} \cdot \frac{z_{R,REF}^*(\lambda, k)}{z_{R,REF}^*(\lambda, k)} \right\} \\
&= \frac{1}{\omega_k} \mathbb{E} \left\{ \angle \frac{z_{L,REF}(\lambda, k) \cdot z_{R,REF}^*(\lambda, k)}{|z_{R,REF}(\lambda, k)|^2} \right\} \\
&= \frac{1}{\omega_k} \mathbb{E} \left\{ \angle z_L(\lambda, k) z_R^*(\lambda, k) \right\} \\
&\cong \frac{1}{\omega_k} \angle \mathbb{E} \left\{ z_L(\lambda, k) z_R^*(\lambda, k) \right\} \\
&= \frac{1}{\omega_k} \angle \mathbf{w}_L^H(k) \mathbf{\Phi}_{vv}(k) \mathbf{w}_R(k).
\end{aligned} \tag{57}$$

Aproximando-se $ITD_{out}^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k))$ e $ITD_{in}^v(\lambda, k)$ por suas médias, desprezando-se as correlações entre numerado e denominador, substituindo-se em (53) e negligenciando-se o fator de escala, após um conjunto de manipulações chega-se na solução apresentada em DOULO *et al.* (2005b):

$$J_{ITD}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = 1 - \cos(\angle \mathbf{w}_L^H(k) \mathbf{\Phi}_{vv}(k) \mathbf{w}_R(k) - \angle \mathbf{q}_L^T \mathbf{\Phi}_{vv}(k) \mathbf{q}_R). \tag{58}$$

De acordo com a teoria duplex, essa função custo deve ser aplicada nos bins de frequência que estão abaixo da frequência de 1,5 kHz, ou seja, preservando a ITD da estrutura fina.

3.4.5 MWF-ITF (*Interaural Transfer Function*)

Uma das técnicas mais exploradas na literatura para redução de ruído com preservação do cenário acústico foi desenvolvida com o intuito de preservar a função de transferência interaural (ITF) da fonte de ruído (BOGAERT *et al.*, 2007).

A função custo total é definida como

$$J_{TOTAL}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = J_{MWF}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) + \xi J_{ITF}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)). \tag{59}$$

A função custo para preservação da função de transferência interaural entre a fonte de ruído e o usuário do aparelho auditivo é definida como

$$\begin{aligned} J_{ITF}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= \mathbb{E} \left\{ \left| \Delta ITF^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \right|^2 \right\} \\ &= \mathbb{E} \left\{ \left| ITF_{out}^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) - ITF_{in}^v(\lambda, k) \right|^2 \right\}. \end{aligned} \quad (60)$$

Os valores instantâneos da função de transferência interaural do caminho acústico entre a fonte de ruído e os microfones de referência foram definidos em (KLASEN *et al.*, 2006) como

$$ITF_{in}^v(\lambda, k) = \frac{v_{L,REF}(\lambda, k)}{v_{R,REF}(\lambda, k)} \quad (61)$$

e da função de transferência interaural de saída do ruído como

$$ITF_{out}^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = \frac{z_{L,REF}(\lambda, k)}{z_{R,REF}(\lambda, k)}. \quad (62)$$

No mesmo trabalho os valores médios da ITF de entrada são definidos como

$$\begin{aligned} ITF_{in}^v(k) &= \mathbb{E} \left\{ ITF_{in}^v(\lambda, k) \right\} \\ &= \mathbb{E} \left\{ \frac{v_{L,REF}(\lambda, k)}{v_{R,REF}(\lambda, k)} \right\} \\ &= \mathbb{E} \left\{ \frac{v_{L,REF}(\lambda, k) v_{R,REF}^*(\lambda, k)}{v_{R,REF}(\lambda, k) v_{R,REF}^*(\lambda, k)} \right\} \\ &= \mathbb{E} \left\{ \frac{\mathbf{q}_L^T \mathbf{v}(\lambda, k) \mathbf{v}^H(\lambda, k) \mathbf{q}_R}{\mathbf{q}_R^T \mathbf{v}(\lambda, k) \mathbf{v}^H(\lambda, k) \mathbf{q}_R} \right\} \\ &\cong \frac{\mathbb{E} \left\{ \mathbf{q}_L^T \mathbf{v}(\lambda, k) \mathbf{v}^H(\lambda, k) \mathbf{q}_R \right\}}{\mathbb{E} \left\{ \mathbf{q}_R^T \mathbf{v}(\lambda, k) \mathbf{v}^H(\lambda, k) \mathbf{q}_R \right\}} = \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\mathbf{q}_R^T \Phi_{vv}(k) \mathbf{q}_R} \end{aligned} \quad (63)$$

e do valor médio da ITF de saída do ruído como

$$\begin{aligned}
ITF_{\text{out}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= \mathbb{E} \left\{ ITF_{\text{out}}^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \right\} \\
&= \mathbb{E} \left\{ \frac{z_{L,\text{REF}}(\lambda, k)}{z_{R,\text{REF}}(\lambda, k)} \right\} \\
&= \mathbb{E} \left\{ \frac{z_{L,\text{REF}}(\lambda, k) z_{R,\text{REF}}^*(\lambda, k)}{z_{R,\text{REF}}(\lambda, k) z_{R,\text{REF}}^*(\lambda, k)} \right\} \\
&\stackrel{(64)}{=} \frac{\mathbb{E} \left\{ z_{L,\text{REF}}(\lambda, k) z_{R,\text{REF}}^*(\lambda, k) \right\}}{\mathbb{E} \left\{ z_{R,\text{REF}}(\lambda, k) z_{R,\text{REF}}^*(\lambda, k) \right\}} \\
&\stackrel{(64)}{=} \frac{\mathbf{w}_L^H(k) \Phi_{\text{vv}}(k) \mathbf{w}_R(k)}{\mathbf{w}_R^H(k) \Phi_{\text{vv}}(k) \mathbf{w}_R(k)} \\
&\stackrel{(64)}{=} \frac{\mathbf{w}_L^H(k) \Phi_{\text{vv}}(k) \mathbf{w}_L(k)}{\mathbf{w}_R^H(k) \Phi_{\text{vv}}(k) \mathbf{w}_L(k)}.
\end{aligned}$$

Sua origem baseia-se na consideração que o cenário acústico é composto por apenas uma fonte pontual de ruído, de forma que podemos expressar os sinais recebidos através de suas HRTFs, como

$$v_{l,m}(\lambda, k) = h_{l,m}(k)v(\lambda, k) \quad (65)$$

em que $h_{l,m}(k)$ é a HRTF, para a frequência discreta k , entre a fonte de ruído e o m -ésimo microfone do lado $l = \{ L, R \}$ (esquerdo/direito). Dessa forma, temos

$$\Phi_{\text{vv}}(k) = \sigma_v^2(k) \mathbf{h}(k) \mathbf{h}^H(k) \quad (66)$$

em que $\sigma_v^2(k) = \mathbb{E} \{ v(\lambda, k) v^*(\lambda, k) \}$ e $\mathbf{h}(k) = [h_{L,1}(k), \dots, h_{L,M}(k), h_{R,1}(k), \dots, h_{R,M}(k)]^T$ é o vetor de HRTFs entre a fonte de ruído e cada um dos microfones. Substituindo (66) em (63) e (64) e o resultado em (60), obtém-se

$$J_{ITF}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = \left| \frac{\mathbf{w}_L^H(k) \mathbf{h}(k) \mathbf{h}^H(k) \mathbf{w}_R(k)}{|\mathbf{w}_R^H(k) \mathbf{h}(k)|^2} - \frac{h_{L,\text{REF}}(k) h_{R,\text{REF}}^*(k)}{|h_{R,\text{REF}}(k)|^2} \right|^2 \quad (67)$$

em que $h_{l,REF}(k)$ é a HRTF, do bin k , entre a fonte de ruído e o microfone frontal $l = \{ L, R \}$ (esquerdo/direito).

3.4.6 MWF-RTF (*Relative Transfer Function*)

Visando controlar as pistas binaurais da componente residual do ruído, foi proposto em (MARQUARDT *et al.*, 2015) a adição de uma restrição linear à função custo do MWF que visa a preservar a ITF instantânea do ruído

$$\min_{\mathbf{w}} J_{\text{MWF}}(\mathbf{w}) \text{ sujeito a } \frac{\mathbf{w}_L^H \mathbf{b}}{\mathbf{w}_R^H \mathbf{b}} = \frac{B_L}{B_R}. \quad (68)$$

em que \mathbf{w} é um vetor composto pelos vetores de coeficientes do aparelho auditivo esquerdo (\mathbf{w}_L) e direito (\mathbf{w}_R) empilhados. O vetor \mathbf{b} é um vetor contendo as funções de transferência acústicas (ATF – do inglês *Acoustic Transfer Function*) entre os microfones de referência do aparelho auditivo esquerdo (B_L) e direito (B_R) e a fonte de ruído e definido como

$$\mathbf{b} = \begin{bmatrix} B_L \\ B_R \end{bmatrix}. \quad (69)$$

A RTF entre os microfones de referência e os aparelhos auditivos é definida como a razão entre as funções de transferência acústica (ATF)

$$RTF_{\text{in}}^v = \frac{ATF_L^v}{ATF_R^v} = \frac{B_L}{B_R} \quad (70)$$

enquanto que a RTF processada da fonte de ruído é definida como a razão entre as componentes filtradas nos aparelhos auditivos esquerdo e direito

$$RTF_{\text{out}}^v = \frac{\mathbf{w}_L^H \mathbf{b}}{\mathbf{w}_R^H \mathbf{b}}. \quad (71)$$

A restrição descrita em (68) pode ser rescrita como

$$\mathbf{w}^H \mathbf{c} = 0, \quad (72)$$

com

$$\mathbf{c} = \begin{bmatrix} \mathbf{b} \\ \varepsilon \mathbf{b} \end{bmatrix}, \quad \varepsilon = -\frac{B_L}{B_R} = -RTF_{in}^v. \quad (73)$$

Usando o método de multiplicadores de Lagrange é possível resolver o problema de otimização com restrição como

$$\mathbf{W}_{MWF-RTF} = \mathbf{R}^{-1} \mathbf{r}_x - \frac{\mathbf{R}^{-1} \mathbf{c} \mathbf{c}^H \mathbf{R}^{-1} \mathbf{r}_x}{\mathbf{c}^H \mathbf{R}^{-1} \mathbf{c}} \quad (74)$$

onde

$$\mathbf{R} = \begin{bmatrix} \Phi_{yy} & 0_{2M} \\ 0_{2M} & \Phi_{yy} \end{bmatrix}, \quad \mathbf{r}_x = \begin{bmatrix} \phi_{xL} \\ \phi_{xR} \end{bmatrix} \quad (75)$$

em que o vetor \mathbf{r}_x contém os vetores de correlação cruzada da componente de fala em todos microfones com os microfones de referência do canal esquerdo (ϕ_{xL}) e direito (ϕ_{xR}).

Embora seja uma única técnica que apresente uma solução fechada, até o presente momento, nenhum trabalho apresentou experimentos psicoacústicos de lateralização.

3.4.7 Quadro comparativo entre as técnicas apresentadas

A Tabela 4 mostra um resumo das principais técnicas de redução de ruído baseadas no MWF com suas principais vantagens e desvantagens.

Tabela 4 – Quadro comparativo entre diferentes variações da técnica de MWF.

Técnica	Vantagens	Desvantagens
MWF		
SDW	<ul style="list-style-type: none"> ▪ Alta redução de ruído ▪ Preserva apenas a fala 	<ul style="list-style-type: none"> ▪ Fala e ruído na mesma direção
μ	<ul style="list-style-type: none"> ▪ Preserva posição da fala e ruído; ▪ Fácil ajuste de parâmetro. 	<ul style="list-style-type: none"> ▪ Baixa redução de ruído;
ITF	<ul style="list-style-type: none"> ▪ Notáveis resultados psicoacústicos. 	<ul style="list-style-type: none"> ▪ Baixa redução de ruído; ▪ Dificil ajuste de parâmetro;
RTF	<ul style="list-style-type: none"> ▪ Possui solução fechada 	<ul style="list-style-type: none"> ▪ Complexa estimação das ATFs; ▪ Não foram apresentados resultados psicoacústicos.
ITD+ILD	<p>Comparado com MWF-ITF:</p> <ul style="list-style-type: none"> ▪ Resultados psicoacústicos semelhantes; ▪ Redução de ruído superior; ▪ Melhor inteligibilidade; ▪ Maior conforto acústico. 	<ul style="list-style-type: none"> ▪ Necessita ajuste de dois parâmetros.

4 MATERIAIS E MÉTODOS

Neste capítulo é apresentado o conjunto de materiais e métodos utilizados nos experimentos realizados, assim como a estratégia desenvolvida para lidar com o problema de pesquisa apresentado na Seção 1.2. A Figura 21 apresenta um diagrama que visa a exemplificar e facilitar a compreensão do método experimental empregado no trabalho. Cada bloco especifica uma das etapas envolvidas no processo.

O primeiro bloco representa a geração da cena acústica virtual utilizada para os experimentos. Essa etapa consiste na escolha dos arquivos que representarão as fontes de fala e de ruído para a geração da cena acústica. Em sequência, essas fontes serão posicionadas (virtualmente) e os sinais destinados para cada microfone serão gerados respeitando a posição de cada fonte.

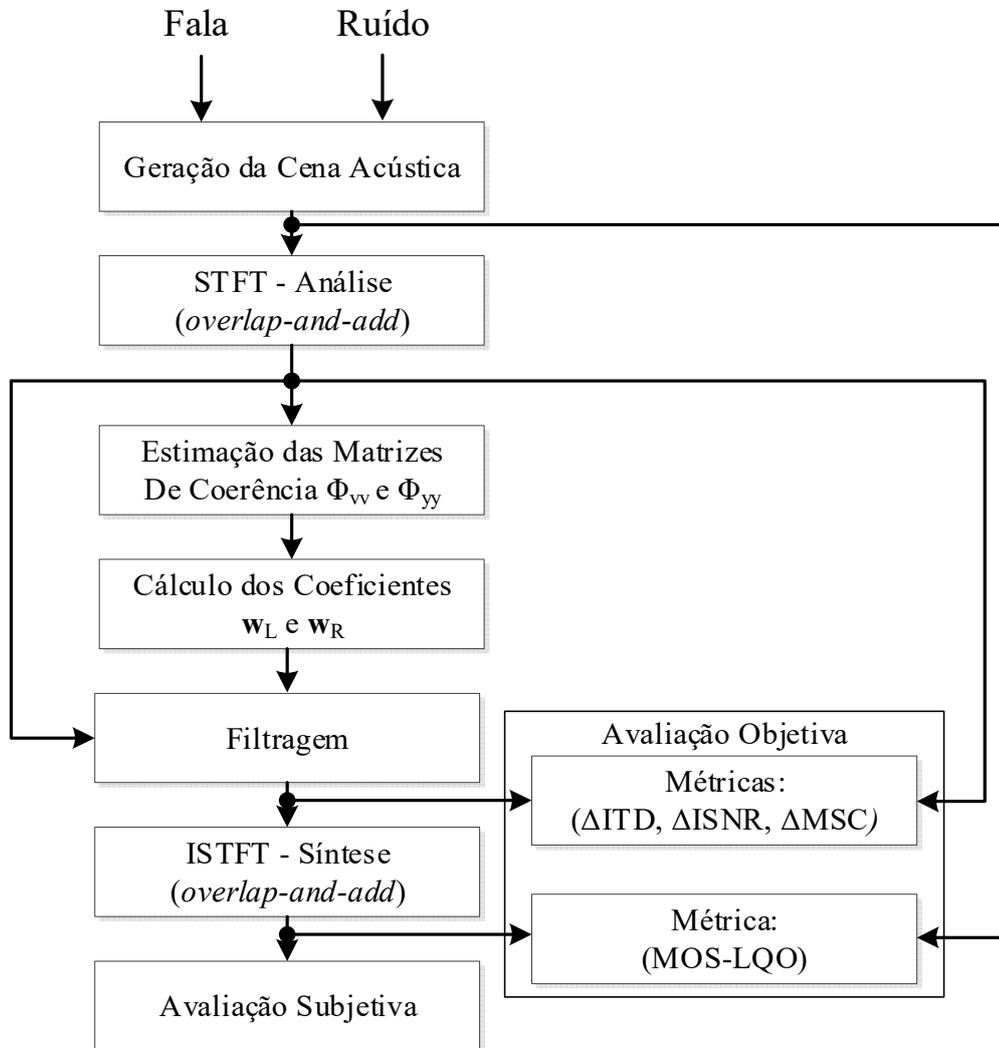
Com todos os sinais de entrada, em cada um dos microfones disponíveis, a transformação para o domínio da frequência é realizada através da STFT. Em sequência, com os sinais no domínio da frequência, as matrizes de coerência dos sinais de entrada (microfones) Φ_{yy} e do ruído Φ_{vv} são estimadas com o auxílio de um detector de fala². Além disso, essas matrizes de coerência são estimadas com o conhecimento completo dos sinais que serão processados, dado que estamos buscando pelo máximo desempenho de todos métodos avaliados. A seguir, os coeficientes do filtro são calculados através da minimização das funções custo apresentadas no Capítulo 3, utilizando um algoritmo de otimização *quasi-newton*. Essas três etapas são efetuadas *off-line*, ou seja, feitas previamente à etapa de filtragem e assumindo-se o conhecimento individualizado do sinal de fala e de ruído.

O próximo passo na sequência é a filtragem dos sinais pelos coeficientes w_L e w_R previamente calculados. A cada trecho em que a filtragem é realizada, algumas métricas como ΔITD , $\Delta iwSNR$ e $\Delta IILD$ são calculadas. Os sinais filtrados são transformados de volta para o domínio do tempo através do método *weighted-overlap-and-add* (WOLA) (CROCHIERE, 1980). São gerados então três arquivos de áudio de saída, um apenas com a fala filtrada, outro apenas com ruído filtrado e um arquivo completo contendo a fala contaminada processada. A avaliação objetiva de qualidade com a métrica PESQ é efetuada com o arquivo contendo apenas

² Nesse trabalho o detector de fala é considerado muito próximo do ideal, sendo obtido através da análise e marcação manual dos sinais de fala envolvidos.

a fala contaminada processada. Por fim a avaliação psicoacústica de localização é feita com os arquivos contendo apenas ruído.

Figura 21 – Metodologia experimental utilizada no trabalho.

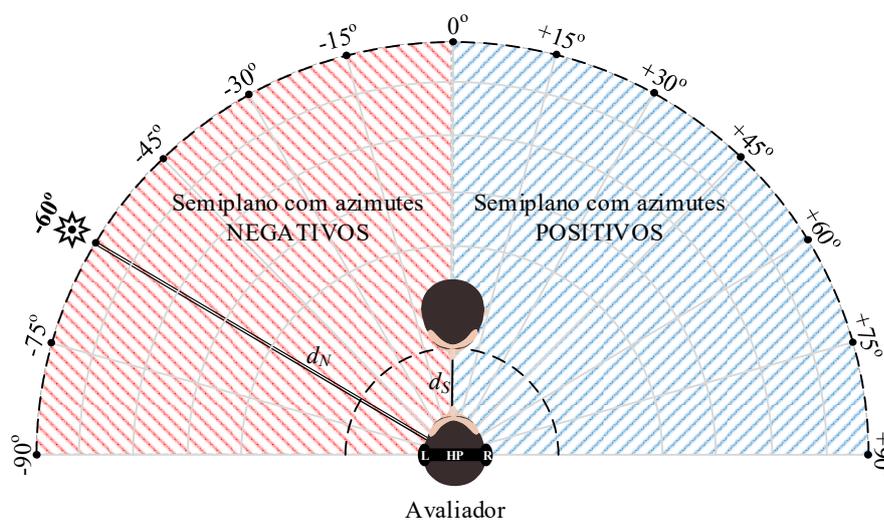


4.1 GERAÇÃO DO CENÁRIO ACÚSTICO

O objetivo geral do trabalho consiste em desenvolver uma forma de preservar a lateralização original de fontes sonoras direcionais através da ITD. Por esse motivo, cenas acústicas compostas por fontes em diferentes azimutes devem ser geradas para os experimentos. Nesses experimentos as fontes sonoras que compõem a cena acústica devem ser espacialmente posicionadas em seus respectivos azimutes. A Figura 22 apresenta um diagrama exemplificando a cena acústica virtual usada nos experimentos. Experimentos de localização

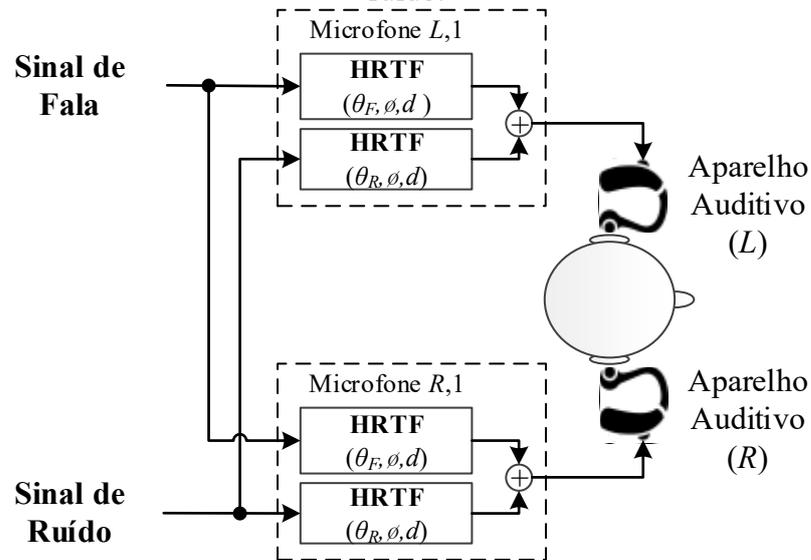
em campo livre demandam considerável infraestrutura como câmara anecoica, alto-falantes, amplificadores, entre outros. Por isso, experimentos de lateralização com fones de ouvido se tornam uma opção mais barata e com resultados confiáveis. Todos experimentos realizados no presente trabalho consideram apenas uma fonte de fala e uma fonte de ruído. A fonte de fala é mantida no azimuth de 0° e a fonte de ruído pode alternar entre os hemisférios, apresentando azimutes positivos ou negativos.

Figura 22 – Cenário acústico virtual gerado para os experimentos.



As localizações das fontes sonoras são emuladas através da filtragem dos sinais por funções transferências que mapeiam um determinado caminho acústico entre a fonte sonora e cada um dos microfones. As HRTFs (*Head Related Transfer Function*) são previamente estimadas, dentro de uma câmara anecoica, com o auxílio de manequins constituídos por cabeça e o torso, semelhantes aos de um homem adulto, para diferentes distâncias (d), elevação (ϕ) e azimutes (θ). No presente trabalho as HRTFs foram obtidas do banco de dados de KAYSER *et al.* (2009). Essa base de dados disponibiliza HRTFs estimadas a 80cm ou 3m de distância do voluntário, elevação -5° , 0° e $+5^\circ$ e azimutes entre -90° e $+90^\circ$ em espaços de 5° .

Figura 23 – Exemplo de geração de cenário acústico com uma fonte de fala e uma fonte de ruído.



A Figura 23 apresenta um esquema sobre o processo de geração dos arquivos de áudio para um sistema de aparelhos binaurais contendo apenas um microfone em cada aparelho. Nesse caso são utilizadas duas HRTFs para posicionar a fonte de ruído no azimute (θ_R) e outras duas para a fonte de fala no azimute (θ_F).

Partiu-se da premissa que o usuário de aparelhos auditivos está sempre de frente para a fonte de fala dentro de uma situação de diálogo no cenário acústico. A nomenclatura utilizada para caracterizar um determinado cenário acústico composto por uma fonte de fala e uma fonte de ruído seguirá o padrão S_0N_θ , significando: fala (S) com azimute de 0° (fixa) e ruído (N) com azimute θ° .

4.2 AVALIAÇÃO DA FUNÇÃO CUSTO J_{ITD}

Nesta seção são apresentados dois experimentos com a função custo mostrada em (52). Ambos têm como objetivo o entendimento dos motivos pelos quais a ITD, na forma utilizada, não é capaz de preservar a lateralização da fonte de ruído direcional.

O primeiro experimento consistiu na aplicação da função custo (52) dentro de uma cena acústica S_0N_{-60} , e no segundo experimento o cenário acústico S_0N_{60} . Cada experimento foi composto por 50 realizações, cada uma consistindo na aplicação dos coeficientes ótimos para 128 valores diferentes de β variando entre 10^{-12} e 10^9 . A Figura 24 apresenta os resultados

da métrica ΔITD do ruído no primeiro experimento e a Figura 25 o ΔITD do ruído no segundo experimento.

Figura 24 – Comportamento da métrica ΔITD do ruído em 50 realizações para a configuração

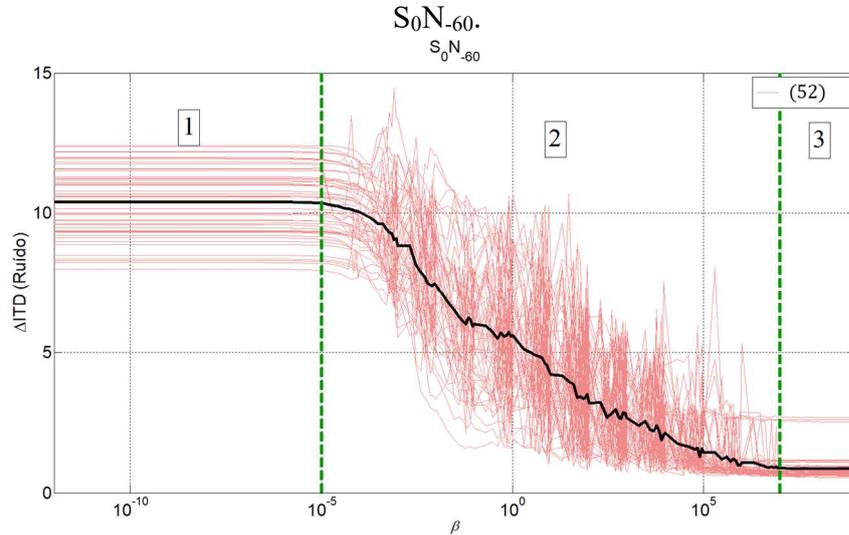
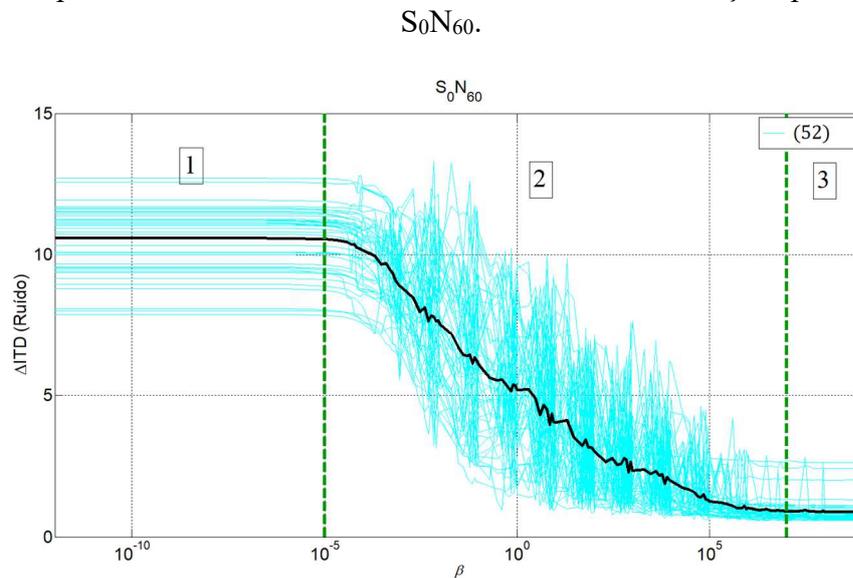


Figura 25 – Comportamento da métrica ΔITD do ruído em 50 realizações para a configuração



Na Figura 24 e Figura 25, a linha preta contínua em ambas as figuras representa o valor médio da métrica de ΔITD do ruído para as 50 realizações em cada experimento. Ambas as figuras são propositalmente separadas por duas linhas verticais tracejadas em verde que servem para destacar que existem basicamente três regiões no comportamento da métrica ΔITD à medida que β varia. A região chamada de Wiener (indicada pelo número 1) ocorre quando

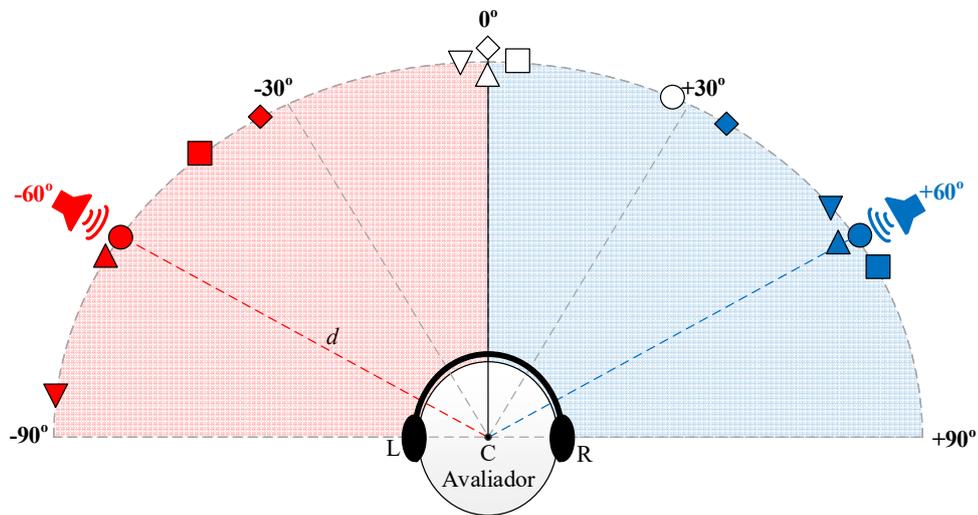
valores muito pequenos de β (inferiores a 10^{-5}) são utilizados. Nesse intervalo, J_{MWF} se sobrepõe dentro da função custo total apresentada em (52). Dentro dessa região são atingidos os maiores níveis de redução de ruído, porém devido aos altos valores de ΔITD , não há preservação da lateralidade da fonte de ruído em ambos cenários acústicos. Quando β assume valores num intervalo entre aproximadamente 10^{-5} e 10^7 , diz-se que a métrica se encontra na região de transição (região indicada na Figura pelo número 2). Nessa região há uma alta variabilidade entre as realizações tornando difícil a escolha de um β de confiança que garanta um baixo ΔITD . Por fim, quando β ultrapassa o valor de aproximadamente 10^7 , a métrica estabiliza na grande maioria das realizações, por isso a região foi chamada de estabilização (região indicada pelo número 3). Nessa região não há redução de ruído e ΔITD é reduzido naturalmente.

À primeira vista o raciocínio parece lógico. Para preservar a lateralização da fonte de ruído com a função custo apresentada em (52), bastaria escolher um valor de β da ordem de 10^7 ou superior. Entretanto, outras métricas além da ΔITD devem ser avaliadas paralelamente, principalmente as que envolvem a fala (sinal de interesse). Em outras palavras, a escolha de um β elevado, apesar de manter a sensação original da localização da fonte de ruído, compromete a qualidade e a inteligibilidade da fala. Os resultados de conforto, qualidade e confiabilidade mostrados na Tabela 1, foram obtidos com (52) para $\beta = 10^7$ mostrando que embora o ΔITD seja pequeno essas outras métricas são comprometidas.

Por fim, alguns experimentos psicoacústicos também foram efetuados com essa função custo para verificar a efetividade da preservação da lateralidade da fonte do ruído, corroborando a expectativa de que baixos valores de ΔITD são capazes de garantir a preservação da lateralização da fonte de ruído. Os experimentos foram realizados de acordo com a metodologia apresentada na Seção 2.6.3 e Apêndice A.

O primeiro experimento consistiu na verificação da capacidade de 5 voluntários em identificar a lateralização do ruído não processado nos azimutes de -60° , 0° e $+60^\circ$. Previamente ao experimento, foi informado aos voluntários que as posições das fontes poderiam estar em qualquer azimute entre -90° e $+90^\circ$. Cada voluntário escutou duas vezes o ruído em cada azimute, em ordem aleatória e informou o azimute identificado. Os resultados médios para cada voluntário são apresentados na Figura 26, indicando que os 5 voluntários foram capazes de identificar o azimute original do ruído não processado.

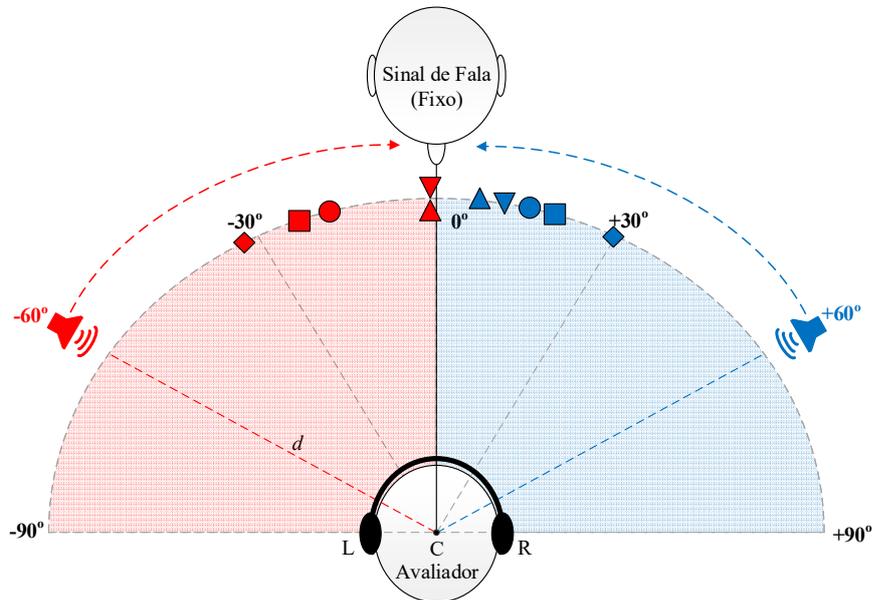
Figura 26 – Resultados psicoacústicos médios com arquivos de áudio do ruído não processado. Voluntários: 1 (o), 2 (Δ), 3 (∇), 4 (\square), 5 (\diamond). Resultados referentes a -60° em vermelho, 0° em branco e $+60^\circ$ em azul.



Os voluntários 1 e 2 se destacaram por acertar o valor exato em ambas as tentativas. O voluntário 1 acertou os azimutes $+60^\circ$ e -60° enquanto que o voluntário 2 acertou os azimutes 0° e $+60^\circ$. De maneira geral, os voluntários se mostraram capazes de identificar o azimute de chegada do sinal não processado.

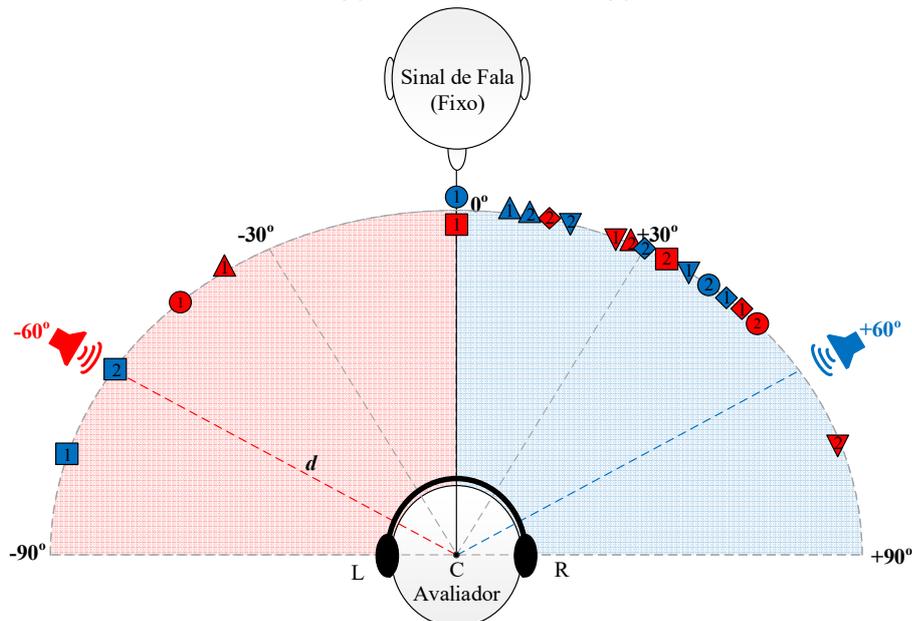
O segundo experimento subjetivo foi realizado com o áudio processado por (52) com um $\beta = 0$. Dessa forma, apenas a função custo J_{MWF} estará atuando. O ruído processado é separado do sinal de fala, permitindo que o voluntário vote no azimute percebido escutando apenas o ruído residual (evitando influência do efeito de precedência do sinal de fala). Duas importantes considerações podem ser feitas baseadas nos resultados dos experimentos psicoacústicos apresentados na Figura 27. Em primeiro lugar, o filtro de Wiener não gerou inversão de hemisfério. Em segundo lugar, é possível notar uma tendência dos resultados em direção ao azimute de 0° , onde se encontra a fonte de fala. Esse efeito já havia sido comentado na Seção 3.3. Praticamente todos os resultados estão concentrados entre os azimutes de $+30^\circ$ e -30° .

Figura 27 – Resultados psicoacústicos médios obtidos para o ruído processado pela solução ótima da equação (52) para $\beta = 0$ (J_{MWF}). Voluntários: 1 (o), 2 (Δ), 3 (∇), 4 (\square), 5 (\diamond). Resultados referentes a -60° em vermelho e $+60^\circ$ em azul.



Em sequência, foi avaliada a sensação subjetiva dos voluntários para o ruído processado com (52) para $\beta = 10^7$ (dentro da chamada região de estabilização). A Figura 28 apresenta as votações individuais de todos os voluntários (ao invés de suas médias). A partir dos resultados obtidos verifica-se que todos os voluntários incorreram em pelo menos uma inversão de hemisfério em relação ao azimute original da fonte de ruído.

Figura 28 – Resultados psicoacústicos médios obtidos para o ruído processado pela solução ótima da equação (52) para $\beta = 10^7$. Voluntários: 1 (o), 2 (Δ), 3 (∇), 4 (\square), 5 (\diamond). Resultados referentes a -60° em vermelho e $+60^\circ$ em azul.



Em resumo, a avaliação preliminar de preservação da ITD no método MWF multicanal não apresentou resultados objetivos e psicoacústicos satisfatórios. A obtenção de uma baixa ΔITD requer valores muito altos para a variável β fazendo com que o sinal de fala seja afetado em qualidade, inteligibilidade e conforto acústico. Mesmo para valores de β extremamente altos, os voluntários incorreram não apenas em erros elevados de azimute como também em inversões de hemisfério.

4.3 ANÁLISE DO ESTIMADOR DE ITD

Como visto no capítulo anterior, a ITD medida em segundos é definida como a fase da razão entre as componentes de uma determinada fonte sonora nos microfones de referência (*Left* e *Right*) dos aparelhos auditivos. A preservação da *ITD* da fonte de ruído no domínio da frequência é alcançada através da preservação da *IPD* em cada bin até a frequência de 1,5 kHz. Sendo assim, a *ITD* pode ser definida dentro dessa banda como

$$\begin{aligned} ITD_{in}^v(\lambda, k) &= \frac{1}{2\pi f_s k} IPD_{in}^v(\lambda, k) \\ &= \frac{1}{2\pi f_s k} \left(\angle \frac{v_{L,REF}(\lambda, k)}{v_{R,REF}(\lambda, k)} + 2\pi p(k) \right) \end{aligned} \quad (76)$$

em que \angle indica a fase do argumento e f_s e frequência de amostragem. O inteiro $p(k)$ é o fator de desenrolamento (*unwrapping*) de fase, que é desconhecido *a priori*, por isso o ângulo da fase é calculado módulo 2π . Isso torna a fase ambígua acima de 1500 Hz devido ao tamanho e formato da cabeça humana. Para frequências acima de 1500 Hz, $p(k)$ pode ser considerado zero (RASPAUD, VISTE e EVANGELISTA, 2010). Sob tais condições o ruído médio de entrada *IPD* (em radianos), para uma dada janela de tempo, pode ser calculado através das seguintes aproximações

$$\begin{aligned}
IPD_{in}^v(k) &= \mathbb{E} \left\{ \angle \frac{v_{L,REF}(\lambda, k)}{v_{R,REF}(\lambda, k)} \times \frac{v_{R,REF}^*(\lambda, k)}{v_{R,REF}^*(\lambda, k)} \right\} \\
&= \mathbb{E} \left\{ \angle v_{L,REF}(\lambda, k) v_{R,REF}^*(\lambda, k) \right\} \\
&\cong \angle \mathbb{E} \left\{ v_{L,REF}(\lambda, k) v_{R,REF}^*(\lambda, k) \right\} \\
&= \angle \mathbf{q}_L^T \mathbf{\Phi}_{vv}(k) \mathbf{q}_R
\end{aligned} \tag{77}$$

em que $\mathbf{\Phi}_{vv}(k) = \mathbb{E} \{ \mathbf{v}(\lambda, k) \mathbf{v}^H(\lambda, k) \}$. Usando a mesma abordagem para $z_L(\lambda, k)$ and $z_R(\lambda, k)$ resulta o valor médio do IPD do ruído na saída

$$IPD_{out}^v(k) = \angle \mathbf{w}_L^T(k) \mathbf{\Phi}_{vv}(k) \mathbf{w}_R(k) \tag{78}$$

que é definida como a diferença de fase entre os ruídos de saída em ambos alto-falantes. Usando as equações (77) e (78) acima chega-se na função custo da *ITD* em função de *IPD*

$$\begin{aligned}
J_{ITD}^v(k) &= |ITD_{out}^v(k) - ITD_{in}^v(k)|^2 \\
&= \frac{1}{4\pi^2 f_s^2 k^2} |IPD_{out}^v(k) - IPD_{in}^v(k)|^2 \\
&= \frac{1}{4\pi^2 f_s^2 k^2} J_{IPD}^v(k).
\end{aligned} \tag{79}$$

Na prática, a estimação da *ITD* é composta por vários estimadores da diferença de fase entre os microfones de referência do aparelho esquerdo e direito para cada bin. Resultando na função custo mostrada em (52), na qual $(2\pi f_s k)^{-2}$ foi incluído em $\beta(k)$.

$$J_{TOTAL}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = J_{MWF}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) + \beta(k) \cdot J_{IPD}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \tag{80}$$

Através do teorema do limite central, é razoável assumir, que no domínio STFT, as componentes do ruído de entrada em ambos microfones de referência, $v_{L,REF}$ e $v_{R,REF}$, para um dado bin k e frame λ são variáveis aleatórias complexas de média zero com distribuição normal e matriz de coerência dada por

$$\Phi_{v_L v_R} = \begin{bmatrix} \sigma_{v_L, \text{REF}}^2 & \rho \sigma_{v_L, \text{REF}} \sigma_{v_R, \text{REF}} \\ \rho^* \sigma_{v_L, \text{REF}} \sigma_{v_R, \text{REF}} & \sigma_{v_R, \text{REF}}^2 \end{bmatrix} \quad (81)$$

em que $\sigma_{v_L}^2 = \mathbb{E} \{ |v_{L, \text{ref}}|^2 \}$, $\sigma_{v_R}^2 = \mathbb{E} \{ |v_{R, \text{ref}}|^2 \}$, e $\rho = \mathbb{E} \{ v_{L, \text{ref}} v_{R, \text{ref}}^* \} / (\sigma_{v_L} \sigma_{v_R})$ é o coeficiente complexo de coerência dos microfones de referência dos aparelhos direito e esquerdo.

Considerando a função de densidade de probabilidade da razão de duas variáveis aleatórias correlacionadas, circularmente simétricas, com distribuição normal:

$$\psi = \psi_r + j\psi_i = \frac{v_{L, \text{REF}}}{v_{R, \text{REF}}}, \quad (82)$$

assumindo média zero e coeficiente de correlação complexo $\rho = \rho_r + j\rho_i$, a PDF de ψ em coordenadas retangulares é definida em YAN e REN (2016) como

$$p_{\psi_r, \psi_i}(\psi_r, \psi_i) = \frac{1 - |\rho|^2}{\pi \sigma_{v_L, \text{REF}}^2 \sigma_{v_R, \text{REF}}^2} \left(\frac{|\psi|^2}{\sigma_{v_L, \text{REF}}^2} + \frac{1}{\sigma_{v_R, \text{REF}}^2} - 2 \frac{\rho_r \psi_r + \rho_i \psi_i}{\sigma_{v_L, \text{REF}} \sigma_{v_R, \text{REF}}} \right)^{-2}. \quad (83)$$

Aplicando transformação de variáveis para coordenadas polares através de

$$\psi_r = |\psi| \cos \theta, \quad \psi_i = |\psi| \sin \theta \quad (84)$$

em que θ denota a fase de ψ . A PDF conjunta em coordenadas polares é dada por

$$p_{|\psi|, \theta}(|\psi|, \theta) = |\psi| p_{\psi_r, \psi_i}(\psi_r, \psi_i). \quad (85)$$

A PDF marginal da fase é obtida através da integração de (85) em relação a $|\psi|$.

$$p_\theta(\theta) = \int_0^\infty p_{|\psi|, \theta} d|\psi|, \quad (86)$$

resultando em

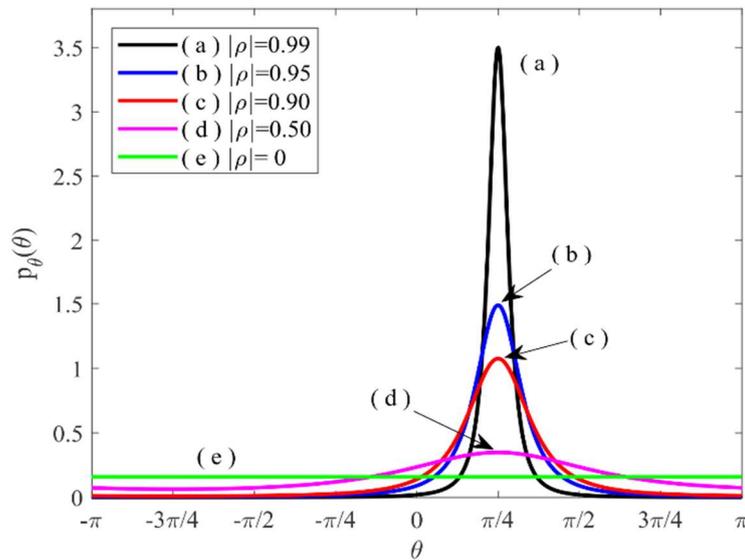
$$p_{\theta}(\theta) = \frac{1-|\rho|^2}{2\pi(1-\eta^2)} \left[\frac{\eta}{\sqrt{1-\eta^2}} \arccos(-\eta) + 1 \right] \quad (87)$$

em que

$$\eta = |\rho| \cos(\angle \rho - \theta). \quad (88)$$

A Figura 29 apresenta através de um exemplo a morfologia de (87) para $\rho = |\rho| \exp(j\pi/4)$ e diferentes valores de $|\rho|$. Pode ser observado que a tendência central permanece fixa em $\pi/4$, porém a dispersão aumenta à medida que o valor absoluto do coeficiente da coerência $|\rho|$ diminui.

Figura 29 – Função de densidade de probabilidade marginal para a fase de $v_{L,\text{ref}}/v_{R,\text{ref}}$ como função de $|\rho| = |\mathbb{E} \{ v_{L,\text{ref}}/v_{R,\text{ref}}^* \}| / (\sigma_{v_L} \sigma_{v_R})$.



Como podemos observar na Figura 29, a dispersão dos resultados aumenta consideravelmente quando $|\rho|$ diminui. No caso extremo onde $|\rho|=0$, no qual não há coerência entre os componentes de ruído residual $z_L(\lambda, k)$ e $z_R(\lambda, k)$, a fase apresenta uma distribuição uniforme. Por isso, conclui-se que o desempenho do estimador depende da magnitude do coeficiente de coerência.

4.4 HIPÓTESE DA COERÊNCIA INTERAURAL

Como discutido na Seção 2.2.3, a preservação da lateralização de fontes sonoras pontuais com frequências abaixo de 1,5 kHz depende da preservação da ITD de estrutura fina (WIGHTMAN e KISTLER, 1992). Entretanto, também há evidências de que a sensibilidade para a ITD varia, dependendo de quão similar as formas de onda são entre as duas orelhas (TRAHIOTIS, BERNSTEIN e AKEROYD, 2001). Uma medida estatística de similaridade é fornecida pela correlação cruzada binaural calculada como uma função do atraso entre as orelhas direita e esquerda. Em condições ideais, dentro de uma câmara anecoica ou campo livre, com a fonte sonora em frente a um indivíduo, a coerência é sempre muito próxima da unidade (valor máximo), não importando a estrutura espectral do som. Entretanto, quando o som que chega às orelhas é degradado por reflexões acústicas geradas pelas superfícies de uma sala, a coerência é reduzida. Quando essa coerência é reduzida de maneira substancial, a ITD tende a perder o significado devido ao fato de que não existem mais formas de onda com características semelhantes que possam ser comparadas para identificar uma ITD (RAKERD e HARTMANN, 2010). O mesmo acontece para a ILD. Em outras palavras, a coerência interaural (IC) determina a confiabilidade das ILDs e ITDs utilizadas no processo de localização (HADAD, DOCLO e GANNOT, 2016).

A coerência interaural do ruído de entrada foi definida em MARQUARDT, HOHMANN e DOCLO (2013) como

$$IC_{in}^v(k) = \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_L \mathbf{q}_R^T \Phi_{vv}(k) \mathbf{q}_R}} \quad (89)$$

e a coerência interaural do ruído de saída como

$$IC_{out}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = \frac{\mathbf{w}_L^H(k) \Phi_{vv}(k) \mathbf{w}_R(k)}{\sqrt{\mathbf{w}_L^H(k) \Phi_{vv}(k) \mathbf{w}_L(k) \mathbf{w}_R^H(k) \Phi_{vv}(k) \mathbf{w}_R(k)}} \quad (90)$$

onde a coerência interaural (IC) é um número complexo, e muitas vezes é difícil a interpretação de seu significado. Entretanto, existe uma medida chamada Magnitude Quadrática da Coerência

(do inglês MSC - *Magnitude Squared Coherence*) que nada mais é do que o módulo da coerência ao quadrado, ou seja, um número real. Com a definição dessa medida surge uma importante métrica de avaliação que mede o grau de degradação da coerência interaural entre os sinais de entrada que chegam aos microfones de referência e os sinais de saída enviados para cada um dos alto-falantes, conhecida como Erro MSC (ΔMSC), e definida como,

$$\Delta MSC^v = \frac{1}{N-1} \sum_{k=1}^{N-1} \left| MSC_{\text{out}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) - MSC_{\text{in}}^v(k) \right| \quad (91)$$

em que N corresponde ao número de bins, e

$$\begin{aligned} MSC_{\text{out}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= |IC_{\text{out}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k))|^2 \\ MSC_{\text{in}}^v(k) &= |IC_{\text{in}}^v(k)|^2. \end{aligned} \quad (92)$$

O resultado dessa métrica varia entre 0 e 1, onde 0 indica que não houve degradação de coerência e 1 que houve total degradação.

Na Figura 30 e Figura 31 são apresentados resultados de simulações do erro MSC para os cenários anteriormente descritos com fontes de ruído nos azimutes de -60° e $+60^\circ$ e processados pela solução ótima apresentada em (52). Em ambos os casos, na região de Wiener, o erro de MSC é relativamente alto (0,48). Isso mostra que o MWF modifica a IC do sinal de saída. Na região de transição, o erro de MSC aumenta, chegando a 0,74 com $\beta = 1$ para os dois azimutes. Quando se utiliza o valor $\beta = 10^7$ o erro de MSC diminui para 0,57, permanecendo relativamente alto. Esse é, provavelmente, o principal motivo dos resultados subjetivos apresentados na Seção 4.2 não terem sido bem-sucedidos. Um alto valor de β é capaz de reduzir significativamente o erro de ITD mas não é capaz de reduzir o valor do erro MSC decorrente do processamento do ruído pelos filtros de Wiener nas orelhas direita e esquerda. Sendo assim a ITD perde o seu significado.

Figura 30 – ΔMSC da técnica apresentada em (52) para a configuração S_0N_{-60} .

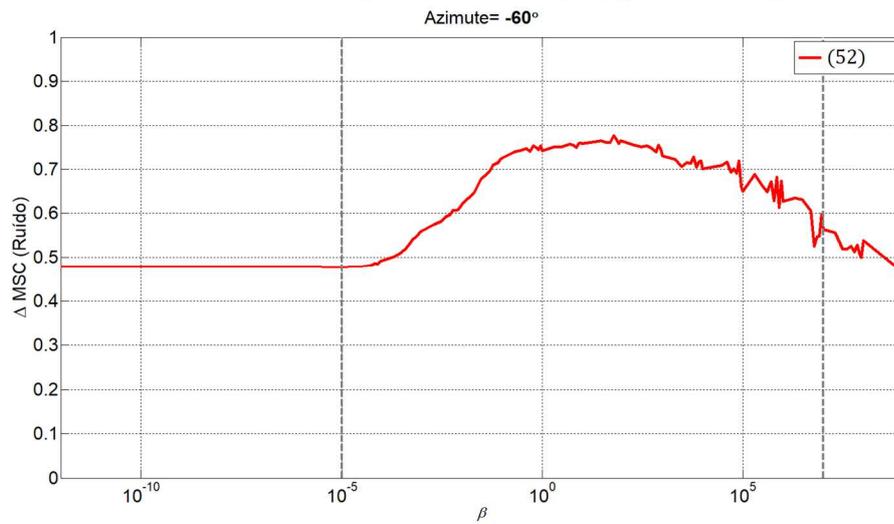
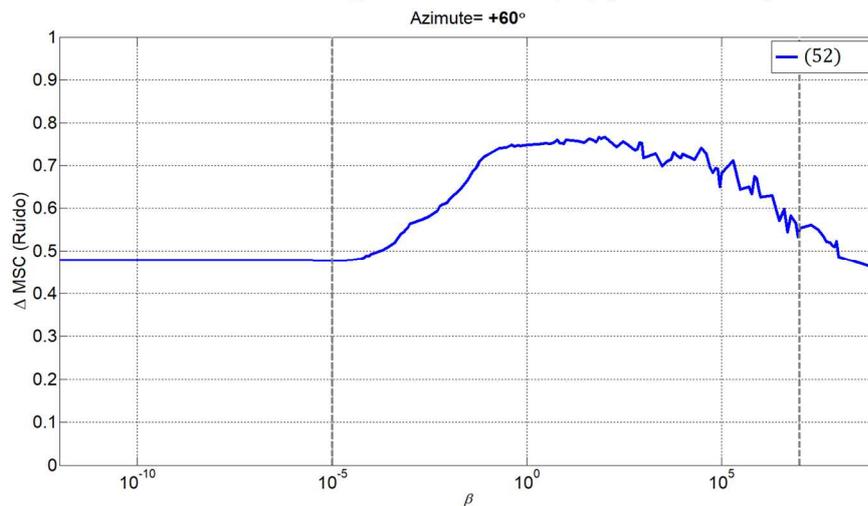


Figura 31 – ΔMSC da técnica apresentada em (52) para a configuração S_0N_{60} .



4.4.1 Função custo da magnitude quadrática

A partir do exposto na seção anterior, há indicações de que para preservar o azimute da fonte de ruído é necessário não apenas minimizar a ITD, como também a MSC. Sendo assim, uma nova função custo pode ser definida visando minimizar o erro MSC usando o mesmo raciocínio aplicado na função custo da ITD, ou seja

$$\begin{aligned}
J_{TOTAL}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= J_{MWF}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \\
&+ \beta J_{IPD}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \\
&+ \gamma J_{MSC}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)).
\end{aligned} \tag{93}$$

A função custo que visa a preservar ΔMSC é definida como

$$\begin{aligned}
J_{MSC}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= |\Delta MSC_v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k))|^2 \\
&= |MSC_{out}^v(\lambda, k, \mathbf{w}_L(k), \mathbf{w}_R(k)) - MSC_{in}^v(\lambda, k)|^2.
\end{aligned} \tag{94}$$

Os valores da MSC do ruído de entrada em cada *frame* são definidos como

$$\begin{aligned}
MSC_{in}^v(\lambda, k) &= |IC_{in}^v(\lambda, k)| \\
&= \left| \frac{v_{L,REF}(\lambda, k)v_{R,REF}^*(\lambda, k)}{\sqrt{(v_{L,REF}(\lambda, k)v_{L,REF}^*(\lambda, k))(v_{R,REF}(\lambda, k)v_{R,REF}^*(\lambda, k))}} \right|.
\end{aligned} \tag{95}$$

e na saída como

$$\begin{aligned}
MSC_{out}^v(\lambda, k) &= |IC_{out}^v(\lambda, k)| \\
&= \left| \frac{z_{L,REF}(\lambda, k)z_{R,REF}^*(\lambda, k)}{\sqrt{(z_{L,REF}(\lambda, k)z_{L,REF}^*(\lambda, k))(z_{R,REF}(\lambda, k)z_{R,REF}^*(\lambda, k))}} \right|.
\end{aligned} \tag{96}$$

A MSC média de entrada pode ser aproximada por

$$\begin{aligned}
MSC_{in}^v(k) &= \mathbb{E}\{MSC_{in}^v(\lambda, k)\} \\
&= \mathbb{E}\left\{ \left| \frac{v_{L,REF}(\lambda, k)v_{R,REF}^*(\lambda, k)}{\sqrt{(v_{L,REF}(\lambda, k)v_{L,REF}^*(\lambda, k))(v_{R,REF}(\lambda, k)v_{R,REF}^*(\lambda, k))}} \right| \right\} \\
&\approx \left| \frac{\mathbb{E}\{v_{L,REF}(\lambda, k)v_{R,REF}^*(\lambda, k)\}}{\sqrt{\mathbb{E}\{v_{L,REF}(\lambda, k)v_{L,REF}^*(\lambda, k)\} \mathbb{E}\{v_{R,REF}(\lambda, k)v_{R,REF}^*(\lambda, k)\}}} \right| \\
&\approx \left| \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_L \cdot \mathbf{q}_R^T \Phi_{vv}(k) \mathbf{q}_R}} \right|
\end{aligned} \tag{97}$$

e da MSC de saída

$$\begin{aligned}
MSC_{out}^v(k) &= \mathbb{E}\{MSC_{out}^v(\lambda, k)\} \\
&= \mathbb{E}\left\{\frac{z_{L,REF}(\lambda, k)z_{R,REF}^*(\lambda, k)}{\sqrt{(z_{L,REF}(\lambda, k)z_{L,REF}^*(\lambda, k))(z_{R,REF}(\lambda, k)z_{R,REF}^*(\lambda, k))}}\right\} \\
&\cong \left|\frac{\mathbb{E}\{z_{L,REF}(\lambda, k)z_{R,REF}^*(\lambda, k)\}}{\sqrt{\mathbb{E}\{z_{L,REF}(\lambda, k)z_{L,REF}^*(\lambda, k)\}\mathbb{E}\{z_{R,REF}(\lambda, k)z_{R,REF}^*(\lambda, k)\}}}\right| \\
&\cong \left|\frac{\mathbf{w}_L^T \Phi_{vv}(k) \mathbf{w}_R}{\sqrt{\mathbf{w}_L^T \Phi_{vv}(k) \mathbf{w}_L \cdot \mathbf{w}_R^T \Phi_{vv}(k) \mathbf{w}_R}}\right|.
\end{aligned} \tag{98}$$

Por fim, a função custo que visa controlar o erro MSC é definida como

$$J_{MSC}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = \left| \frac{\mathbf{w}_L^H(k) \Phi_{vv}(k) \mathbf{w}_R(k)}{\sqrt{\mathbf{w}_L^H(k) \Phi_{vv}(k) \mathbf{w}_L(k) \cdot \mathbf{w}_R^H(k) \Phi_{vv}(k) \mathbf{w}_R(k)}} - \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_L \cdot \mathbf{q}_R^T \Phi_{vv}(k) \mathbf{q}_R}} \right|^2. \tag{99}$$

A função custo apresentada em (93) tem β para controlar a preservação da ITD e γ para controlar a preservação da MSC e foi avaliada nos dois experimentos já mencionados, cujos resultados são apresentadas na Figura 32 e na Figura 33 para dois valores de γ iguais a 0,05 e 1. O parâmetro β foi variado da mesma forma que nos demais experimentos.

Figura 32 – Comparação do ΔMSC entre a técnica da equação (52) (vermelho) com a técnica proposta em (93) com $\gamma = 0,05$ (magenta) e $\gamma = 1$ (verde) para a configuração S_0N_{-60} .

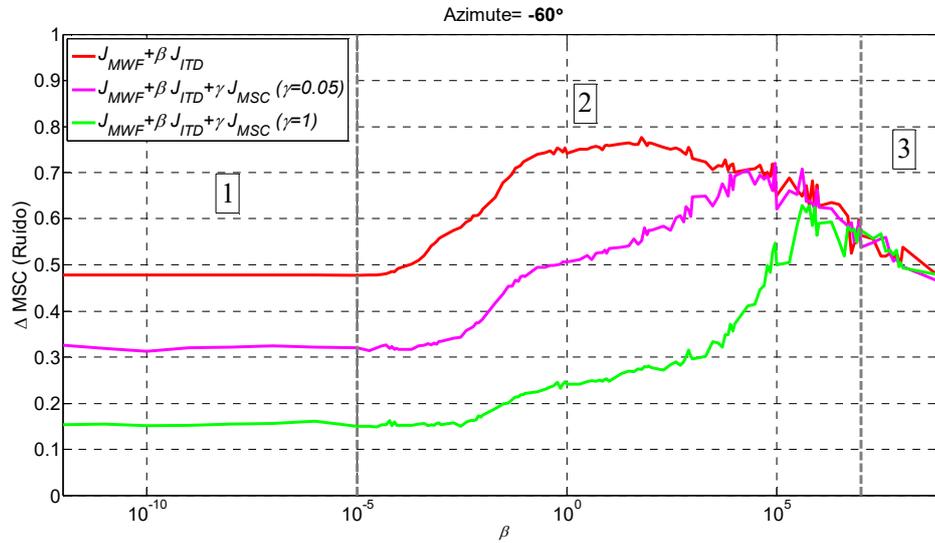
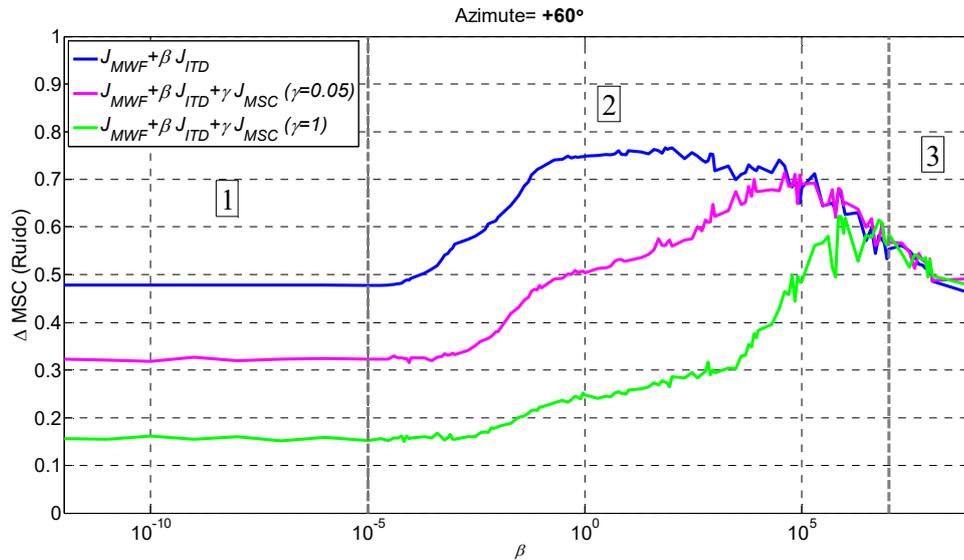


Figura 33 – Comparação do ΔMSC entre a técnica da equação (52) (azimute) com a técnica proposta em (93) com $\gamma = 0,05$ (magenta) e $\gamma = 1$ (verde) para a configuração S_0N_{60} .



Como pode-se observar na Figura 32 e na Figura 33 a solução proposta em (93) reduz o erro MSC nas regiões de Wiener (1) e Transição (2) para os dois valores de γ ($\gamma = 0,05$ e $\gamma = 1$) quando comparada à forma $\gamma = 0$ (equação (52)). Como esperado, a solução com maior γ atingiu os maiores níveis de redução do ΔMSC . Na região de estabilização (3), os valores foram praticamente iguais.

4.4.2 Proposta de uma função custo integrada para preservação da ITD

Nesse mesmo foco de preservação de coerência interaural (IC), uma solução associada com o MWF foi anteriormente proposta em MARQUARDT, HOHMANN e DOCLO (2013). Entretanto, essa solução foi apresentada de forma a preservar a coerência interaural de fontes sonoras difusas e não pontuais. Essas fontes de ruído difusas não possuem uma direção de chegada clara, apresentando uma sensação subjetiva de som espalhado no ambiente. Esse tipo de fonte apresenta valores baixos de IC e a solução visa a, justamente, manter essa baixa coerência. Dessa forma, os trabalhos da área apresentados até esse momento separam claramente as soluções: fontes pontuais de ruído têm sua direcionalidade preservada através da ITF (ITD e ILD subjacentes) enquanto que fontes de ruído difusas tem sua baixa coerência preservada através da IC.

Em MARQUARDT, HOHMANN e DOCLO (2013), foi mostrado que quando se aplica a função custo J_{ITF} em fontes de ruído difusas, o ruído residual passa a apresentar uma sensação subjetiva associada a um determinado azimute de chegada, como se tivesse se tornado uma fonte de ruído pontual. Esse efeito, é obviamente indesejado e deve ser evitado. Entretanto, dentro do conhecimento deste autor, até o momento nenhuma publicação científica ou tecnológica aborda a função custo da IC como método para preservação do azimute de uma fonte de ruído pontual. A função custo J_{IC} é dada por

$$J_{TOTAL}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = J_{MWF}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) + \beta J_{IC}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)). \quad (100)$$

O objetivo de (100) é a preservação da coerência interaural dos sinais que chegam nos microfones de referência. A função custo é definida como

$$\begin{aligned} J_{IC}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= \mathbb{E} \left\{ \left| \Delta IC_v(k, \lambda, \mathbf{w}_L(k), \mathbf{w}_R(k)) \right|^2 \right\} \\ &= \mathbb{E} \left\{ \left| IC_v^{out}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) - IC_v^{in}(k) \right|^2 \right\} \end{aligned} \quad (101)$$

em que IC_v^{in} e IC_v^{out} foram definidas em (89) e (90).

Os valores instantâneos da coerência interaural dos microfones de referência foram definidos em MARQUARDT, HOHMANN e DOCLO (2013) como

$$IC_v^{in}(k, \lambda) = \frac{v_{L,REF}(k, \lambda)v_{R,REF}^*(k, \lambda)}{\sqrt{(v_{L,REF}(k, \lambda)v_{L,REF}^*(k, \lambda))(v_{R,REF}(k, \lambda)v_{R,REF}^*(k, \lambda))}} \quad (102)$$

e da coerência interaural de saída como

$$IC_v^{out}(k, \lambda, \mathbf{w}_L(k), \mathbf{w}_R(k)) = \frac{z_{L,REF}(k, \lambda)z_{R,REF}^*(k, \lambda)}{\sqrt{(z_{L,REF}(k, \lambda)z_{L,REF}^*(k, \lambda))(z_{R,REF}(k, \lambda)z_{R,REF}^*(k, \lambda))}}. \quad (103)$$

No mesmo trabalho, o valor médio da coerência interaural de entrada foi definido como

$$\begin{aligned} IC_v^{in}(k) &= \mathbb{E} \left\{ \frac{v_{L,REF}(\lambda, k)v_{R,REF}^*(\lambda, k)}{\sqrt{|v_{L,REF}(\lambda, k)|^2 \cdot |v_{R,REF}(\lambda, k)|^2}} \right\} \\ &\cong \frac{\mathbb{E} \{ v_{L,REF}(\lambda, k)v_{R,REF}^*(\lambda, k) \}}{\mathbb{E} \left\{ \sqrt{|v_{L,REF}(\lambda, k)|^2 \cdot |v_{R,REF}(\lambda, k)|^2} \right\}} \\ &\cong \frac{\mathbb{E} \{ v_{L,REF}(\lambda, k)v_{R,REF}^*(\lambda, k) \}}{\sqrt{\mathbb{E} \{ |v_{L,REF}(\lambda, k)|^2 \} \cdot \mathbb{E} \{ |v_{R,REF}(\lambda, k)|^2 \}}} \\ &\cong \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_L \cdot \mathbf{q}_R^T \Phi_{vv}(k) \mathbf{q}_R}} \end{aligned} \quad (104)$$

e o valor da coerência interaural de saída definida como

$$\begin{aligned}
IC_v^{out}(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= \mathbb{E} \left\{ \frac{z_{L,REF}(\lambda, k) z_{R,REF}^*(\lambda, k)}{\sqrt{|z_{L,REF}(\lambda, k)|^2 \cdot |z_{R,REF}(\lambda, k)|^2}} \right\} \\
&\cong \frac{\mathbb{E} \{ z_{L,REF}(\lambda, k) z_{R,REF}^*(\lambda, k) \}}{\mathbb{E} \left\{ \sqrt{|z_{L,REF}(\lambda, k)|^2 \cdot |z_{R,REF}(\lambda, k)|^2} \right\}} \\
&\cong \frac{\mathbb{E} \{ z_{L,REF}(\lambda, k) z_{R,REF}^*(\lambda, k) \}}{\left\{ \sqrt{\mathbb{E} \{ |z_{L,REF}(\lambda, k)|^2 \} \cdot \mathbb{E} \{ |z_{R,REF}(\lambda, k)|^2 \}} \right\}} \\
&\cong \frac{\mathbf{w}_L^H(k) \Phi_{vv}(k) \mathbf{w}_R(k)}{\sqrt{\mathbf{w}_L^H(k) \Phi_{vv}(k) \mathbf{w}_L(k) \cdot \mathbf{w}_R^H(k) \Phi_{vv}(k) \mathbf{w}_R(k)}}.
\end{aligned} \tag{105}$$

A função custo final fica

$$J_{IC}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) = \left| \frac{\frac{\mathbf{w}_L^H(k) \Phi_{vv}(k) \mathbf{w}_R(k)}{\sqrt{\mathbf{w}_L^H(k) \Phi_{vv}(k) \mathbf{w}_L(k) \cdot \mathbf{w}_R^H(k) \Phi_{vv}(k) \mathbf{w}_R(k)}}}{\frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_L \cdot \mathbf{q}_R^T \Phi_{vv}(k) \mathbf{q}_R}}} \right|^2. \tag{106}$$

No caso de ruído direcional obtém-se, substituindo-se (66) em (89)

$$\begin{aligned}
IC_{in}^v(k) &= \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_L \mathbf{q}_R^T \Phi_{vv}(k) \mathbf{q}_R}} \\
&= \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \sigma_v^2(k) \mathbf{h}(k) \mathbf{h}^H(k) \mathbf{q}_L \mathbf{q}_R^T \sigma_v^2(k) \mathbf{h}(k) \mathbf{h}^H(k) \mathbf{q}_R}} \\
&= \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \sigma_v^2(k) \mathbf{h}(k) \mathbf{h}^H(k) \mathbf{q}_R \mathbf{q}_R^T \sigma_v^2(k) \mathbf{h}(k) \mathbf{h}^H(k) \mathbf{q}_L}} \\
&= \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R \mathbf{q}_R^T \Phi_{vv}(k) \mathbf{q}_L}} = \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{\sqrt{|\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R|^2}} \\
&= \frac{\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R}{|\mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R|} \\
&= e^{j \angle \mathbf{q}_L^T \Phi_{vv}(k) \mathbf{q}_R + 2\pi p(k)} \\
&= e^{j IPD_{in}^v(k)}
\end{aligned} \tag{107}$$

ou seja, chega-se em uma função de (56). Utilizando-se o mesmo conjunto de manipulações obtém-se

$$\begin{aligned} IC_{\text{out}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &= \frac{\mathbf{w}_L^H(k) \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{w}_R(k)}{|\mathbf{w}_L^H(k) \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{w}_R(k)|} \\ &= e^{jIPD_{\text{out}}^v}. \end{aligned} \quad (108)$$

Sabendo-se que a exponencial complexa pode ser aproximada por série de Taylor como

$$e^{jx} = 1 + jx - \frac{1}{2!}x^2 - j\frac{1}{3!}x^3 + \frac{1}{4!}x^4 + j\frac{1}{5!}x^5 + \dots + \frac{1}{N!}x^N. \quad (109)$$

Utilizando uma aproximação de primeira ordem e substituindo (109) em (107) chega-se em

$$\begin{aligned} IC_{\text{in}}^v(k) &\approx 1 + jITD_{\text{in}}^v(k) \\ IC_{\text{out}}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) &\approx 1 + jITD_{\text{out}}^v(k). \end{aligned} \quad (110)$$

Substituindo (110) em (101) obtém-se

$$J_{IC}^v(k, \mathbf{w}_L(k), \mathbf{w}_R(k)) \Big|_{\text{ruído pontual}} = \left| ITD_{\text{out}}^v(k) - ITD_{\text{in}}^v(k) \right|^2. \quad (111)$$

Isso significa que a função custo descrita em (101) preserva IC para ruídos difusos, como também (dentro das aproximações utilizadas) a ITD para ruídos provenientes de fontes pontuais e, conseqüentemente, reduz o ΔMSC e ΔITD , através do ajuste de um único parâmetro. Na Figura 34 e Figura 35 são apresentados o comportamento do ΔMSC usando (100) em comparação com (52) para os azimutes de -60° e $+60^\circ$ respectivamente.

Figura 34 – ΔMSC para a função custo J_{IC}^v , equação (100) para cenário acústico S_0N_{60} .

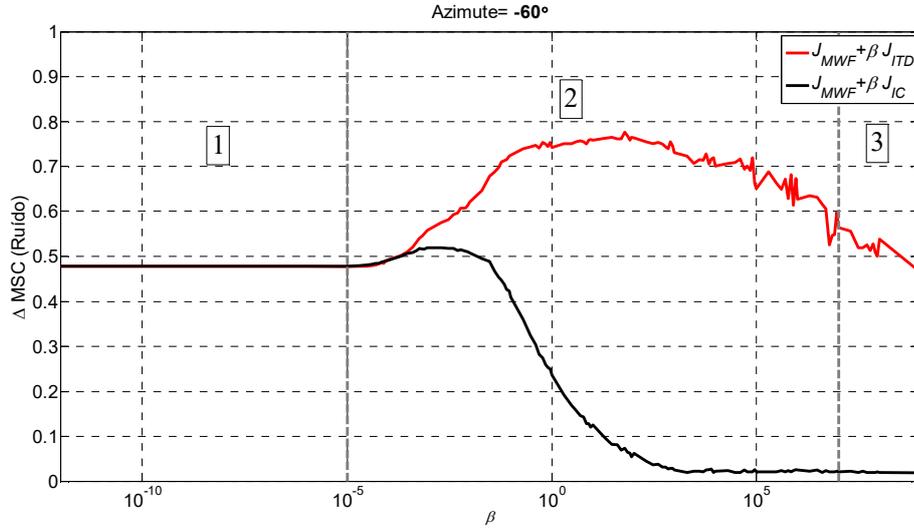
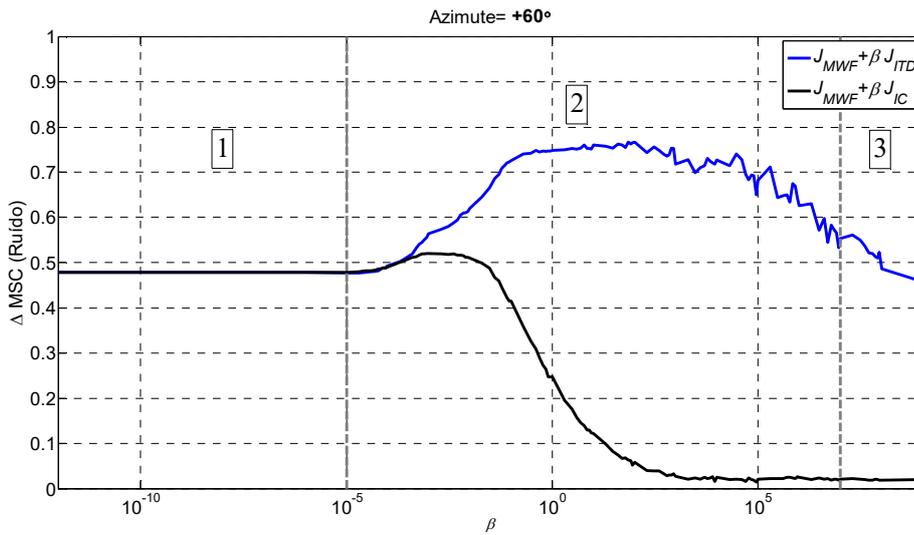


Figura 35 – ΔMSC para a função custo J_{IC}^v , equação (100) para cenário acústico S_0N_{60} .

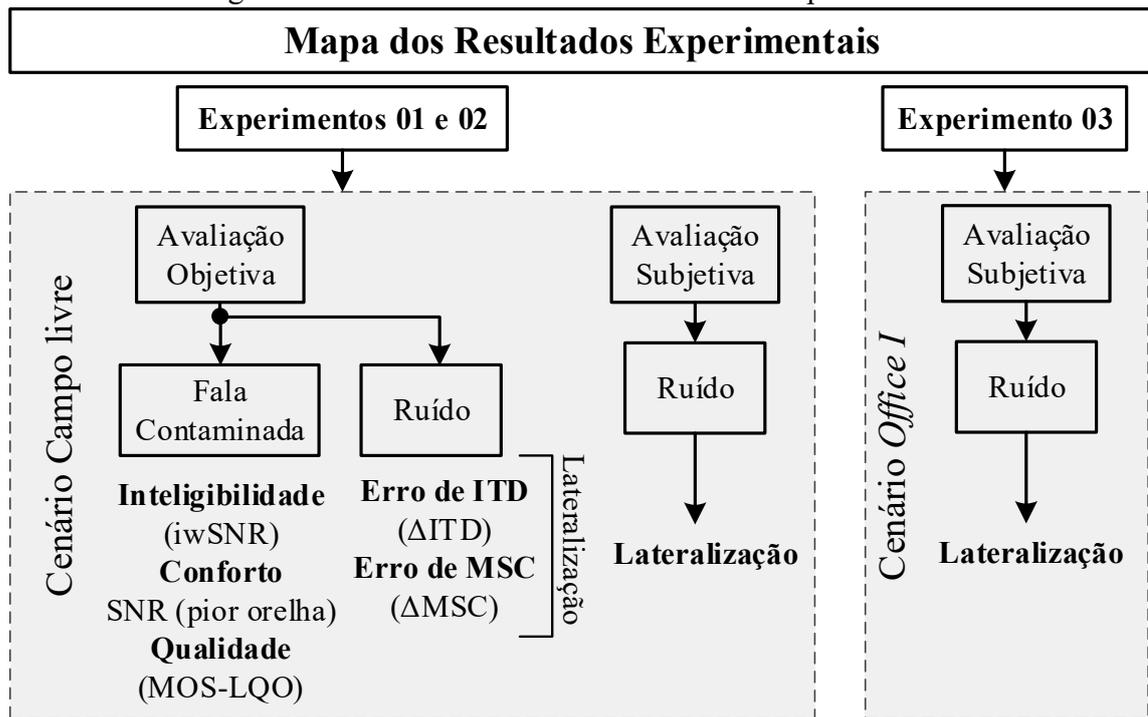


Claramente a função custo J_{IC}^v mostra-se capaz de reduzir o erro MSC à medida que tem seu parâmetro de controle β aumentado.

5 RESULTADO EXPERIMENTAIS

Neste capítulo são apresentados e discutidos os resultados referentes às técnicas apresentadas anteriormente. Os resultados são divididos basicamente em objetivos e psicoacústicos. A avaliação objetiva do ruído é feita com duas métricas voltadas para lateralização da fonte de ruído (Δ ITD e Δ MSC). A avaliação objetiva da fala visa quantificar o impacto na qualidade e inteligibilidade que as técnicas exploradas causam no sinal de interesse. Os experimentos 1 e 2 foram realizados em um cenário de campo livre (sem reverberação) e seguiram o formato mostrado na Figura 36. O experimento 3 foi realizado em um cenário acústico com reverberação (denominado *Office I*) e teve seu desempenho de lateralização avaliado apenas subjetivamente. Mais detalhes sobre cada experimento serão apresentados dentro das respectivas seções.

Figura 36 – Resumo sobre os resultados dos experimentos.

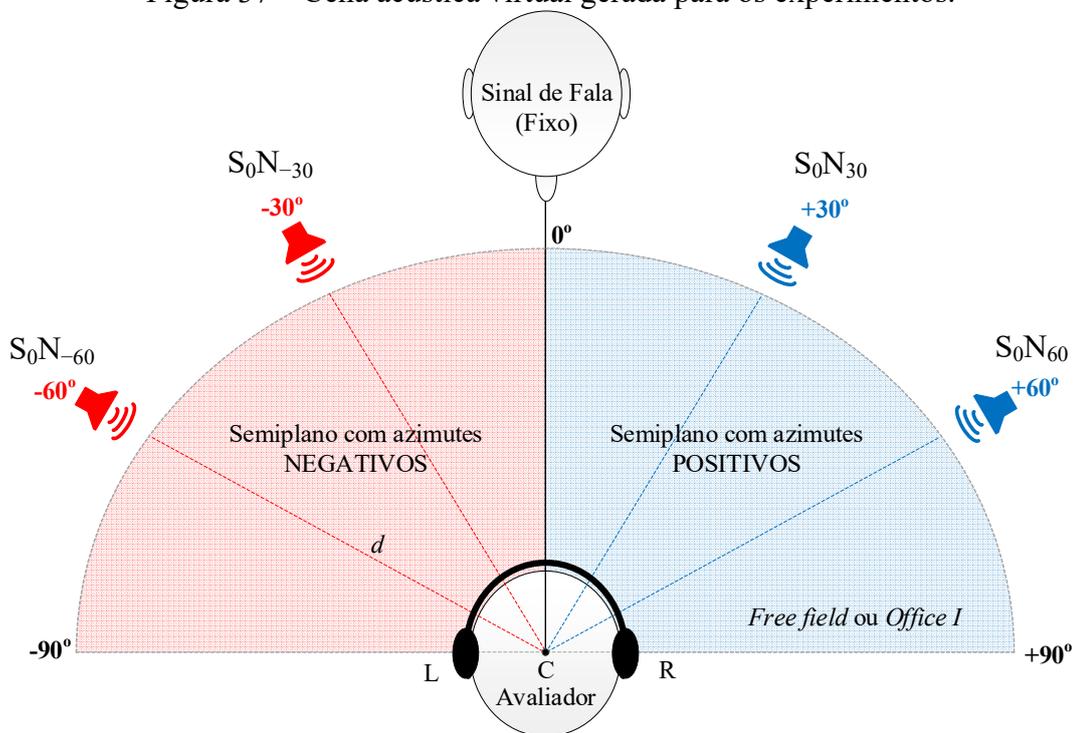


5.1 CENÁRIOS ACÚSTICOS E SIMULAÇÕES

Os cenários acústicos virtuais considerados para a realização de cada experimento são sempre compostos por apenas duas fontes, uma de fala e a outra de ruído, e sua representação é mostrada no diagrama da Figura 37. A fonte de fala está sempre fixada no azimute de 0° ,

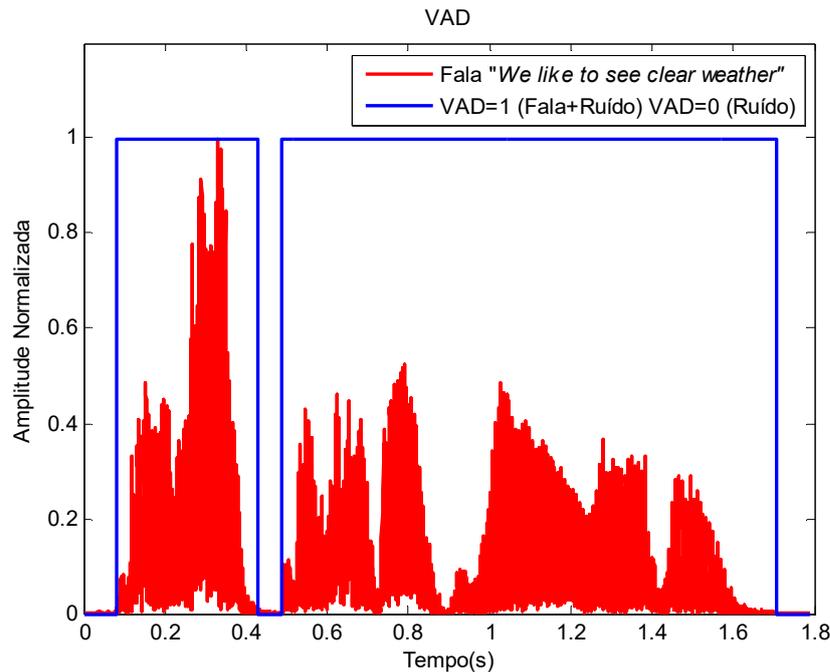
escolha essa suportada pelo fato de usualmente estarmos de frente para o sinal de interesse. A fonte de ruído pode apresentar azimutes de -30° (S_0N_{-30}), -60° (S_0N_{-60}), $+30^\circ$ (S_0N_{30}) ou $+60^\circ$ (S_0N_{60}). Azimutes negativos correspondem ao lado esquerdo do plano sagital do avaliador, enquanto que azimutes positivos correspondem ao lado direito. A distância entre a fonte de fala e o ponto central na cabeça do avaliador (ponto C na Figura 37 é sempre de 80cm, enquanto que entre a fonte de ruído e o mesmo ponto é de 3m. Em todos os experimentos as fontes tem elevação de 0° .

Figura 37 – Cena acústica virtual gerada para os experimentos.



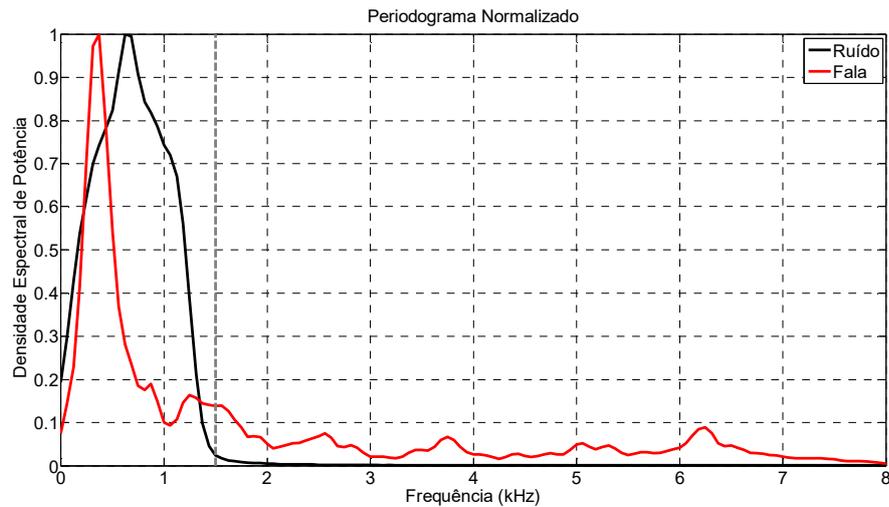
O sinal de fala é uma voz masculina selecionada de ITU-T (1998), contendo uma frase de 2,7 segundos. A voz foi convolvida com uma HRIR de azimute 0° e então aplicado um detector de fala (VAD) obtido através de marcação manual, emulando um VAD ideal, evitando erros de classificação conforme mostrado na Figura 38. Erros de estimação inerentes a um VAD real e seu impacto sobre o desempenho dos métodos de redução de ruído não foram abordados nesse trabalho.

Figura 38 – Classificação dos trechos de voz em fala+ruído (VAD = 1) ou apenas ruído (VAD = 0).



O ruído foi obtido através da filtragem passa-baixas de um ruído branco, limitado em 1,5 kHz (faixa de atuação da ITD de acordo com a teoria duplex) (SANDEL, TEAS, *et al.*, 1955; MILLS, 1960). O ruído foi convolvido com quatro HRIRs diferentes, criando quatro cenários acústicos em campo livre (S_0N_{-30} , S_0N_{-60} , S_0N_{30} e S_0N_{60}) utilizados nos experimentos 1 e 2 e dois cenários acústicos *Office I* (S_0N_{-60} e S_0N_{60}) usados no experimento 3. As HRIRs foram obtidas de uma base de dados binaural multicanal (KAYSER, EWERT, *et al.*, 2009). Nessa base de dados, um manequim com a formato de uma cabeça e torso humanos, utilizando dois aparelhos auditivos *behind-the-ear* compostos por 3 microfones cada ($M_L = M_R = 3$), foi posicionado dentro de uma câmara anecoica (campo livre) e dentro de um escritório (*Office I*). A relação sinal ruído (SNR) do sinal contaminado foi definida como 0 dB na orelha mais próxima da fonte de ruído (chamada de “pior orelha”). Os periodogramas do ruído (preto) e da fala (vermelho) no cenário de campo livre podem ser vistos na Figura 39.

Figura 39 – Periodogramas das fontes de fala (vermelho) e ruído (preto).



A frequência de amostragem foi definida como 16 kHz e os sinais de entrada foram transformados para o domínio da frequência através da transformada de Fourier de tempo curto (STFT) com $N=256$ bins, janela de análise de 128 amostras, zero *padding* e 50% de sobreposição. Os sinais transformados no domínio da STFT foram reconstruídos através do método conhecido como *weighted overlap-and-add* (CROCHIERE, 1980).

As matrizes de coerência $\Phi_{yy}(k)$ e $\Phi_{vv}(k)$ foram calculadas *a-priori*, diretamente dos sinais contaminados com a aplicação do VAD ideal. Isso foi realizado visando evitar possíveis erros nas pistas binaurais devido aos erros de estimação (CORNELIS, MOONEN e WOUTERS, 2011) e para obter o máximo desempenho possível. As matrizes de correlação da fala foram calculadas como $\Phi_{xx}(k) = \Phi_{yy}(k) - \Phi_{vv}(k)$.

Os sinais de fala contaminada, fala e ruído foram processados pelos coeficientes obtidos através da aplicação do método de otimização de Broyden-Fletcher-Goldfarb-Shanno (ARORA, 2004; HABETS e NAYLOR, 2010) para todas as funções custo, assim como pela solução teórica de Wiener (MWF). Os fatores de ponderação definidos para cada técnica são mantidos fixos para todos os bins.

5.2 CENÁRIO ACÚSTICO DE CAMPO LIVRE

Na área de acústica, cenários de campo livre correspondem a situações nas quais não ocorrem reflexões sonoras. Nesses cenários os sons podem ser captados ou percebidos por indivíduos (ou microfones) como se fossem provenientes de fontes sonoras pontuais. Cenários

de campo livre são produzidos artificialmente em câmaras anecoicas. Entretanto, estão presentes em diversas situações cotidianas, como diálogos em ambientes ao ar livre. Por isso, dois experimentos foram realizados nesse tipo de cenário e serão apresentados a seguir.

5.2.1 Experimento 1

Nesse primeiro experimento, o objetivo consiste na avaliação das técnicas apresentadas na Tabela 5 para apenas dois cenários (S_0N_{-60} e S_0N_{60}) em campo livre conforme o esquema apresentado na Figura 37.

Tabela 5 – Resumo das técnicas avaliadas no experimento 1 em campo livre.

Técnica	Sigla	Equação	Proposta pela 1ª vez em	Cor
1	J_{ITD}	(52)	(DOCLO, DONG, <i>et al.</i> , 2005b)	
2a	$J_{MSC}(\gamma = 0,05)$	(93)	Neste trabalho	
2b	$J_{MSC}(\gamma = 1)$			
3	J_{IC}	(100)	Neste trabalho (Fontes Pontuais)	

5.2.1.1 Avaliação Objetiva

Nesta seção os sinais de fala e ruído processados pelas técnicas apresentadas na Tabela 5 são avaliados com métricas objetivas. Primeiramente são apresentados os resultados associados à lateralização da fonte de ruído (ΔITD). Baseados nesses resultados e nos resultados de ΔMSC apresentados no Capítulo 4, são escolhidos os valores do parâmetro de controle β aplicados em cada uma das técnicas. Por fim, cada técnica será analisada em termos de conforto acústico, qualidade e inteligibilidade.

5.2.1.1.1 Erro de ITD (ΔITD) da fonte de ruído

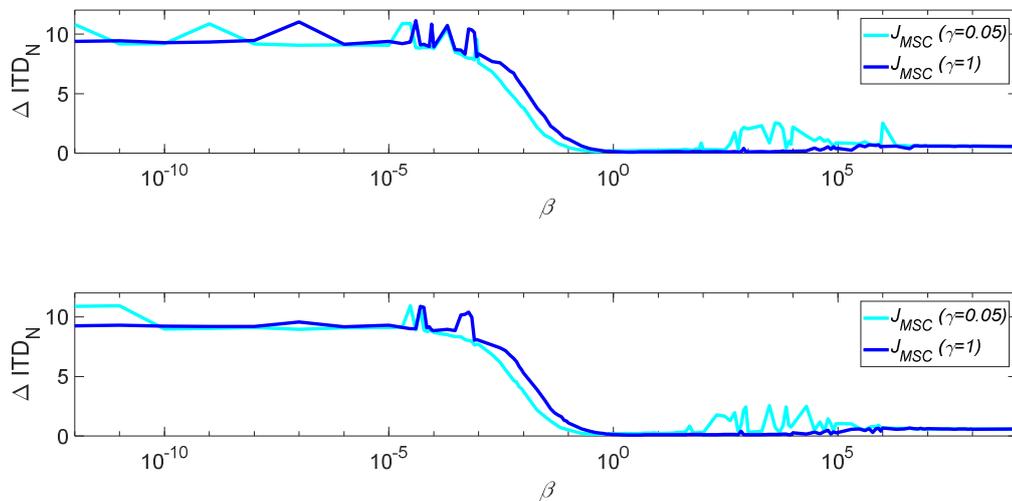
O primeiro resultado apresentado se refere à principal métrica utilizada no processo de lateralização da fonte de ruído pontual, a diferença de ITD entre os sinais que chegam aos microfones de referência (ITD_{in}^v) e os sinais enviados aos alto-falantes (ITD_{out}^v). Os resultados dessa métrica referentes à técnica 1 já foram apresentados na Figura 24 para a configuração

S_0N_{-60} e na Figura 25 para a configuração S_0N_{60} . Experimentos adicionais mostraram que também para outros azimutes da fonte de ruído são necessários altos valores de β para atingir a região de estabilização. Entretanto, como mostrado na Tabela 1, valores de β iguais a 10^7 comprometem significativamente outras métricas como qualidade, inteligibilidade e conforto acústico. Além disso, nessa região os resultados dos experimentos subjetivos apresentados no Capítulo 4, mostraram que essa técnica não preserva a lateralidade da fonte de ruído.

A solução alternativa que visa a reduzir o ΔMSC junto com o ΔITD foi definida na Tabela 5 como técnica 2. Nessa solução, dois parâmetros devem ser controlados, β em relação à função custo da ITD e γ em relação à função custo da MSC. Para a avaliação dessa solução, dois valores fixos de γ foram escolhidos, sendo 0,05 (2a) e 1 (2b), enquanto o parâmetro β é variado conforme os experimentos anteriores.

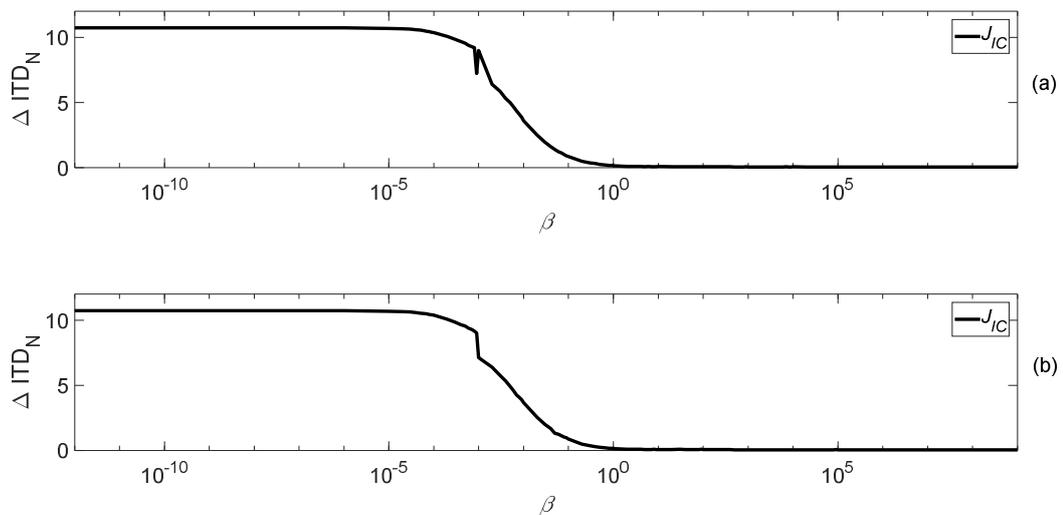
Como pode ser visto na Figura 40, as soluções 2a e 2b geram uma região de estabilização para valores de β menores do que a solução mostrada na Figura 24 e Figura 25. Para ambas soluções, a região de estabilização se inicia para valores de β próximos de 1 (10^0), onde ainda não há um considerável comprometimento das outras métricas de interesse. Outra observação importante diz respeito à modificação do comportamento da métrica na região Wiener, gerando uma flutuação nos valores com β inferiores à 10^{-5} . Essa mudança não preocupa, uma vez que essa região não é utilizada nos casos de preservação da lateralidade da fonte de ruído. Para ambos valores de γ algumas flutuações aparecem para valores de β bem acima de 10^0 .

Figura 40 – Comportamento de ΔITD_N com a variação do parâmetro de controle β da função custo J_{MSC} com dois valores fixos de γ para S_0N_{-60} (a) e S_0N_{60} (b).



Por fim são apresentados os resultados da ΔITD da fonte de ruído através da função custo J_{IC} . Como podemos observar na Figura 41, essa função custo também reduz a região de transição e aumenta a região de estabilização. Além disso não apresenta flutuações significativas na região de estabilização.

Figura 41 – Comportamento de ΔITD com a variação do parâmetro de controle β da função custo J_{IC} para S_0N_{-60} (a) e S_0N_{60} (b).



5.2.1.1.2 Erro de MSC (ΔMSC) da fonte de ruído

Os resultados referentes ao ΔMSC de cada técnica avaliada foram previamente apresentados no Capítulo 4. Em resumo, os resultados mostrados na Figura 30 e Figura 31 confirmaram que a técnica 1 não reduz o ΔMSC mesmo com altos valores de β (da ordem de 10^7), indicando o motivo pelo qual essa técnica não funciona de maneira satisfatória para preservação de fontes sonoras pontuais. As técnicas 2a e 2b diminuíram o ΔMSC inicial na região de *Wiener* (região 1), como apresentado na Figura 32 e Figura 33, entretanto na região de transição (região 2) esse valor aumenta à medida que o parâmetro β também aumenta. Os resultados de ΔMSC apresentados na Figura 34 e Figura 35 indicaram que a técnica 3 é a única capaz de reduzir a ΔMSC progressivamente à medida que β aumenta, da mesma forma que a ΔITD . Por esse motivo, essa técnica será avaliada com dois valores diferentes de β . O valor de $\beta = 1$ apresenta um valor de ΔITD pequeno e estabilizado. O segundo valor de β foi escolhido

visando a atingir os menores valores de ΔMSC e ΔITD simultaneamente. Esse caso será tratado como a máxima lateralização objetiva (M.L.O.) possível.

Para definir esse valor de β , as duas métricas ΔITD e ΔMSC foram apresentadas simultaneamente para o azimute de -60° (Figura 42) e $+60^\circ$ (Figura 43). Para ambos os casos, um $\beta = 10^3$ mostrou-se capaz de atingir valores inferiores a 0,1 nas duas métricas. A Tabela 6 apresenta as técnicas e seus respectivos valores de β encontrados segundo esse critério.

Figura 42 – Função custo J_{IC} . Escolha do menor β que minimiza ΔITD e ΔMSC para o cenário acústico S_0N_{-60} .

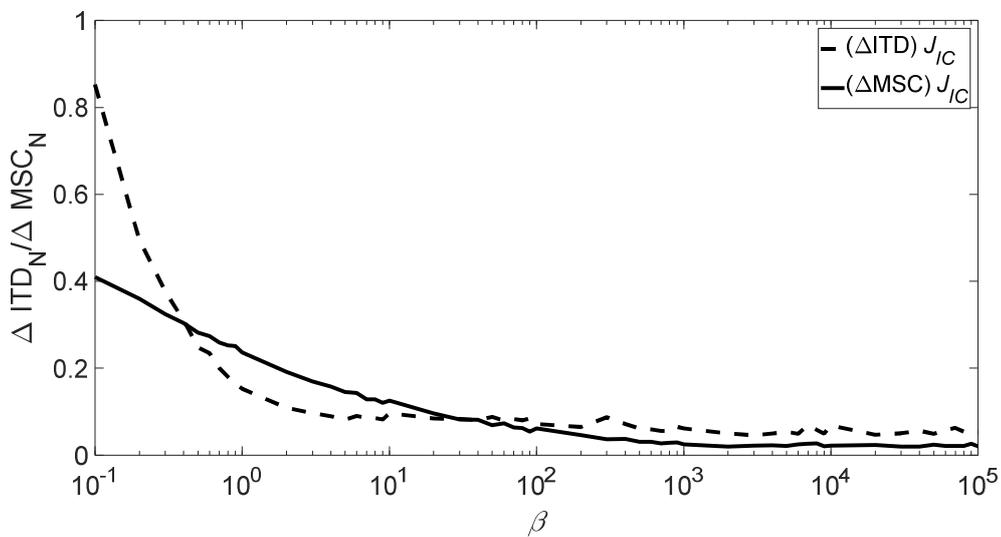


Figura 43 – Função custo J_{IC} . Escolha do menor β que minimiza ΔITD e ΔMSC para o cenário acústico S_0N_{60} .

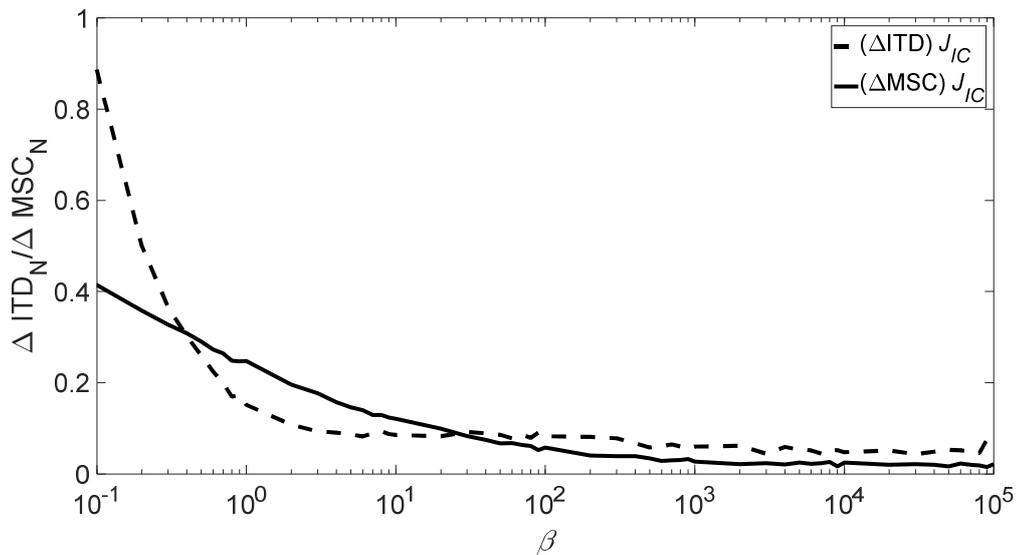


Tabela 6 – Resumo das soluções exploradas no experimento 1.

Técnica	β	Lateralização Objetiva			
		Azimute = -60°		Azimute = $+60^\circ$	
		Δ ITD	Δ MSC	Δ ITD	Δ MSC
1	10^7	0,58	0,57	0,57	0,57
2a	1	0,13	0,51	0,13	0,50
2b	1	0,13	0,24	0,14	0,24
3	1	0,15	0,23	0,15	0,24
3 (M.L.O.)	10^3	0,06	0,02	0,06	0,02

5.2.1.1.3 Conforto acústico

O conforto acústico é uma métrica muito importante pois está diretamente relacionado ao tempo de utilização dos aparelhos auditivos. Esse conforto é medido como a redução de ruído alcançada na chamada “pior orelha”, ou seja, na orelha que está mais próxima da fonte de ruído. Como mencionado no início do capítulo, a SNR inicial é de 0 dB causando um considerável desconforto aos usuários nessa situação. Na Figura 44 (S_0N_{-60}) e na Figura 45 (S_0N_{60}) a linha cinza tracejada refere-se à redução de ruído atingida com o MWF, ou seja, a máxima redução de ruído possível. Em vermelho, como esperado a técnica, 1 apresenta o pior desempenho, com uma SNR de 13,4 dB (S_0N_{-60}) e 15,1 dB (S_0N_{60}). A técnica 2 tem o melhor desempenho com uma SNR de 24,1 dB e 24,3 dB, próxima dos 25,7 dB e 27,2 dB do MWF nos azimutes de -60° e $+60^\circ$ respectivamente. As técnicas 3 e 2b apresentam SNR com valores muito próximos entre si, com uma perda de conforto acústico entre 3,5 dB (S_0N_{-60}) e 5 dB (S_0N_{60}) em relação ao desempenho obtido com o MWF. A solução de máxima lateralização teórica perde muito em redução de ruído, mas com resultado superior ao da técnica 1.

Figura 44 – Resultados de conforto acústico para o cenário acústico S_0N_{-60} .

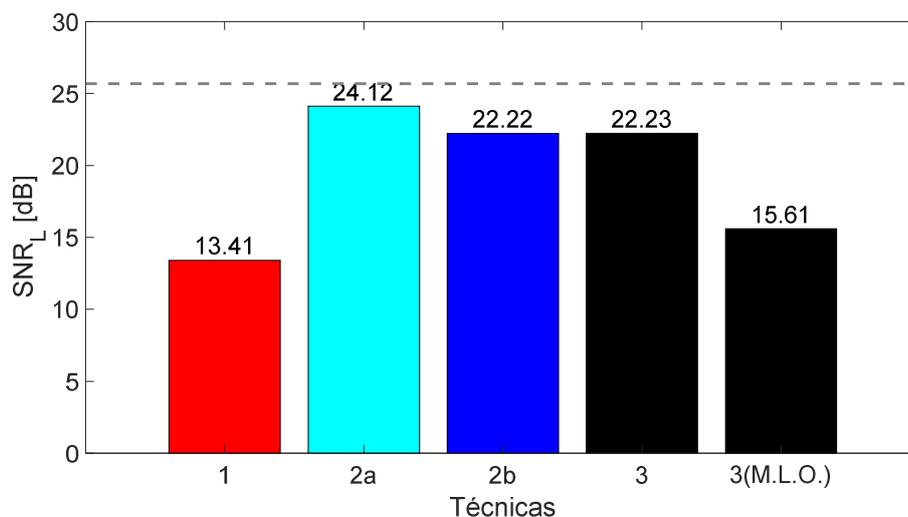
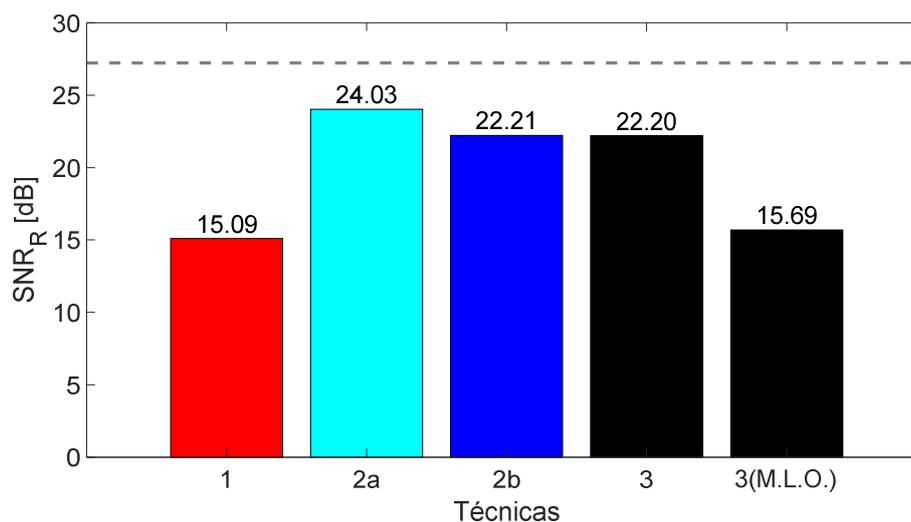


Figura 45 – Resultados de conforto acústico para o cenário S_0N_{60} .



5.2.1.1.4 Qualidade do sinal de fala

Nessa seção são apresentados os resultados referentes à qualidade da fala processada por cada técnica apresentada na Tabela 6. A métrica utilizada foi o MOS-LQO previamente apresentado na Seção 2.5.1. Os resultados de qualidade nos experimentos para o cenário S_0N_{-60} são apresentados na Figura 46 e para o cenário S_0N_{60} na Figura 47.

Figura 46 – MOS-LQO_R referentes às técnicas analisadas para o cenário acústico S₀N₋₆₀.

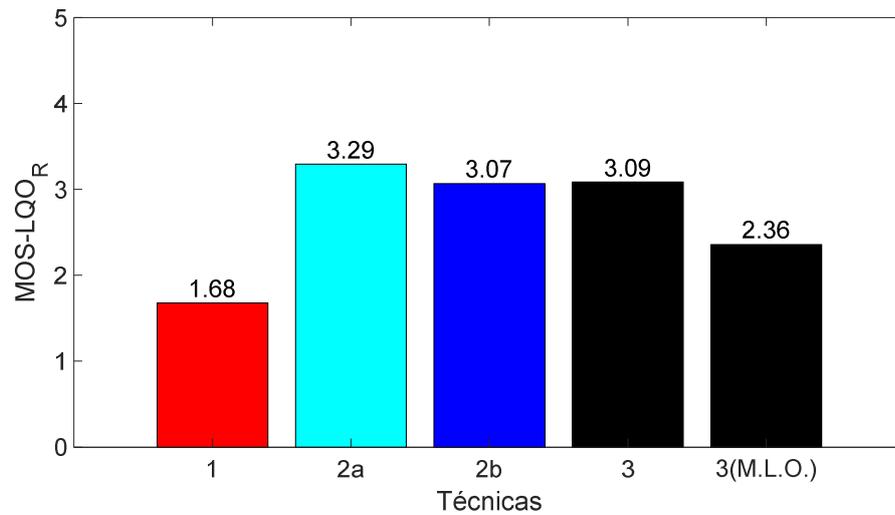
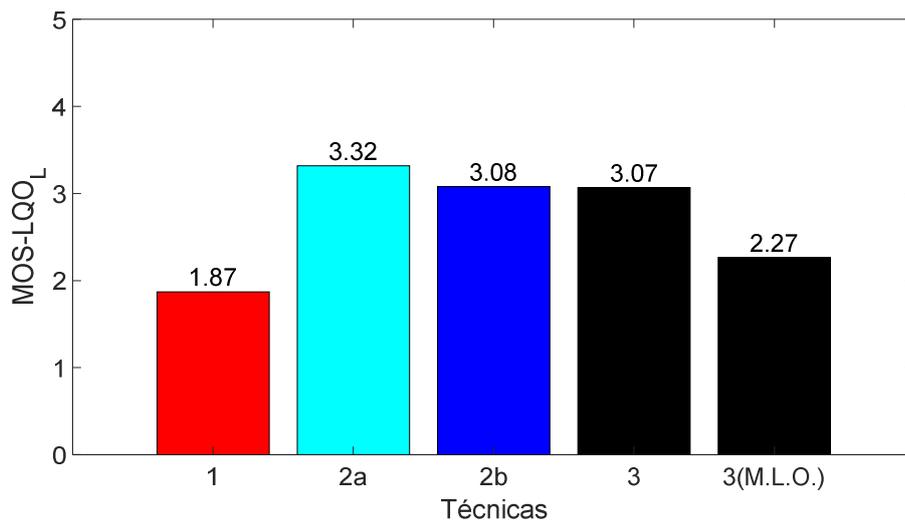


Figura 47 – MOS-LQO_L referentes às técnicas analisadas para o cenário acústico S₀N₆₀.



A Tabela 7 apresenta o MOS-LQO de cada técnica associada com a escala ACR. A técnica 1 apresenta um pequeno valor de MOS-LQO com a fonte de ruído nos dois azimutes, traduzido para a escala ACR como “Ruim”. Isso indica que a fala processada por essa técnica sofre uma considerável e indesejada mudança em suas características. A fala processada pelas técnicas 2a, 2b e 3 recebeu valores de MOS-LQO entre 3 e 4. Por isso, segundo a escala ACR apresentam qualidade “Boa”. Por fim, a solução que visa a alcançar a máxima lateralização 3 (M.L.O.) ficou com o MOS-LQO entre 2 e 3, que equivale ao conceito de qualidade “razoável”.

Tabela 7 – Mapeamento dos resultados para MOS-LQO para ACR.

Técnica	MOS-LQO		ACR
	-60°	+60°	
1	1,7	1,9	Ruim
2a	3,3	3,3	Boa
2b	3,1	3,1	
3	3,1	3,1	
3 (M.L.O.)	2,4	2,3	Razoável

5.2.1.1.5 Inteligibilidade do sinal de fala

A última métrica objetiva apresentada é relacionada com a inteligibilidade da fala, ou seja, à capacidade de identificar e compreender o conteúdo do sinal de fala. A métrica utilizada é o iwSNR (*intelligibility Weighted SNR*), anteriormente apresentada na Seção 2.5.2. Mesmo no caso da técnica 1, ainda é possível a compreensão da fala processada. Os resultados de inteligibilidade do experimento realizado com a fonte de ruído para o cenário S_0N_{-60} são apresentados na Figura 48 e para o cenário S_0N_{60} na Figura 49.

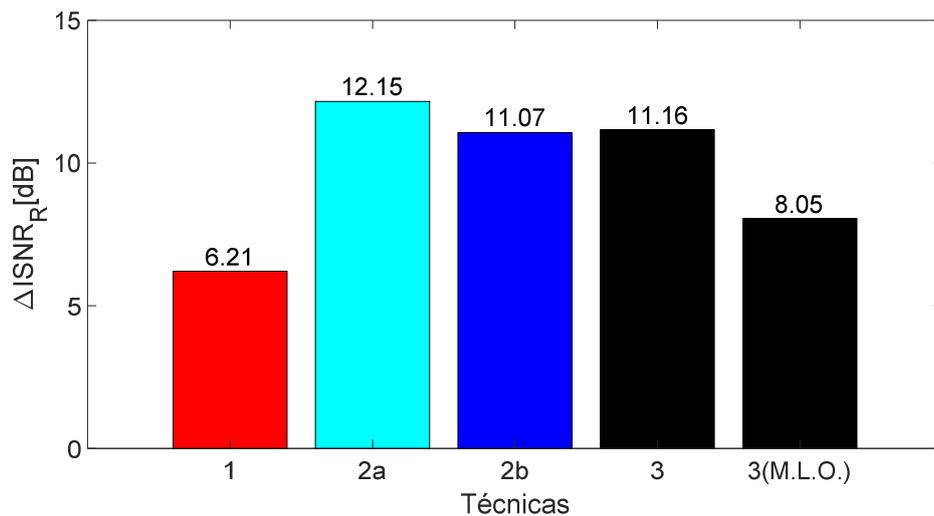
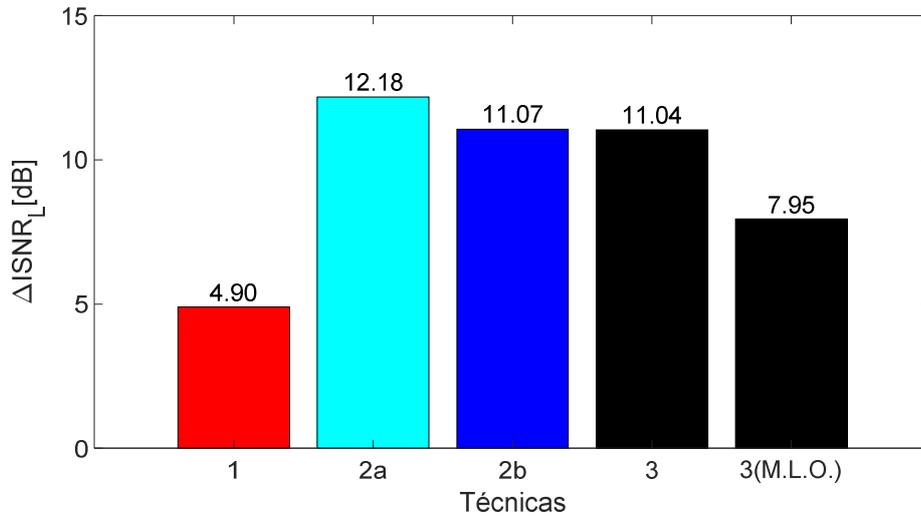
Figura 48 – Resultados de Inteligibilidade das técnicas para o cenário acústico S_0N_{-60} .

Figura 49 – Resultados de Inteligibilidade das técnicas para o cenário acústico S_0N_{60} .



Como esperado a técnica 1 apresenta os piores resultados em termos de inteligibilidade.

5.2.1.2 Avaliação psicoacústica

O objetivo dessa seção é apresentar os resultados dos experimentos psicoacústicos e verificar se os mesmos corroboram os resultados objetivos de lateralização obtidos através das métricas ΔITD e ΔMSC . Foram realizados experimentos baseados em duas cenas acústicas distintas: S_0N_{-60} e S_0N_{60} . O ruído foi processado pelas técnicas apresentadas na Tabela 6. Os cinco voluntários escutaram os arquivos de ruído processado em ordem aleatória com o auxílio de fones de ouvido. O registro dos resultados foi realizado com o auxílio de um programa desenvolvido especificamente para esse fim. Mais detalhes sobre a seleção dos voluntários e sobre o programa podem ser encontrados no Apêndice A.

5.2.1.2.1 Lateralização da fonte de ruído

Os resultados dos experimentos psicoacústicos são mostrados com gráficos de caixa (*boxplots*) na Figura 50 para o cenário acústico S_0N_{-60} e na Figura 51 para o cenário acústico S_0N_{60} . Nesses gráficos as hastes em preto nas extremidades determinam os limites inferiores e superiores, as linhas vermelhas centrais indicam a mediana e as linhas azuis paralelas com a mediana o primeiro e terceiro quartil da votação dos cinco voluntários. A linha contínua em

cinza posicionada no azimute 0° separa os dois semiplanos com azimutes positivos e negativos. A linha contínua em verde indica o azimute real em que a fonte está posicionada originalmente.

Figura 50 – Resultados subjetivos para o cenário acústico S_0N_{-60} .

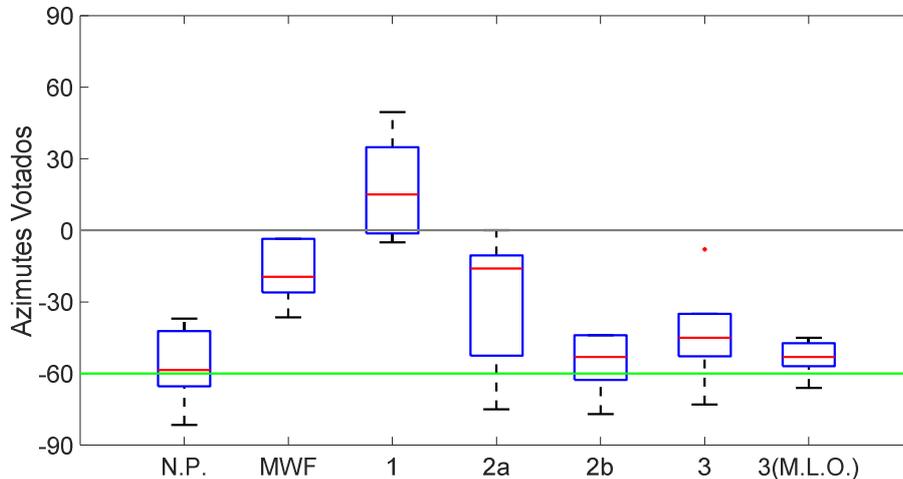
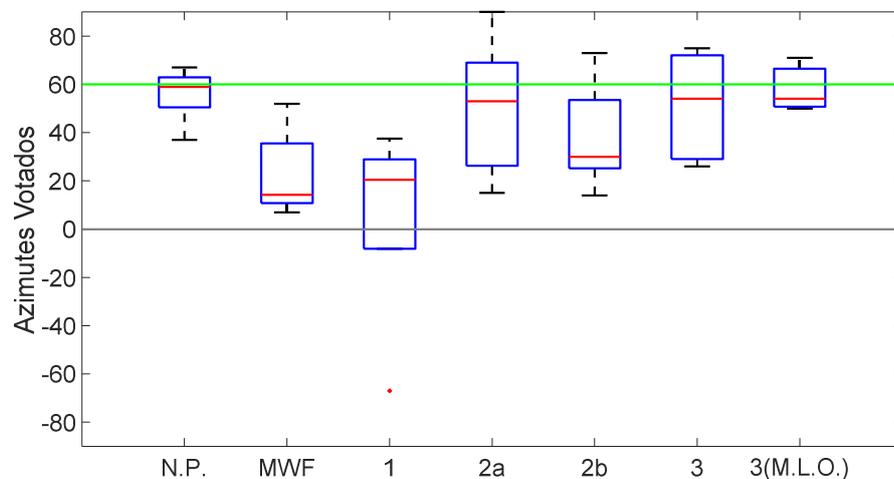


Figura 51 – Resultados subjetivos para o cenário acústico S_0N_{60} .



Os resultados referentes à avaliação do ruído não processado, processado com MWF e com a técnica 1 já haviam sido apresentados de outra forma na Figura 26, Figura 27 e Figura 28 respectivamente. Entretanto a apresentação desses experimentos com gráficos de caixa permite uma comparação entre todas as técnicas em uma única figura para cada azimute.

Para ambos experimentos, os resultados do ruído não processado apresentaram a mediana mais próxima do azimute real da fonte de ruído. Isso enfatiza que o grupo de

voluntários possui uma boa capacidade de lateralizar a fonte de ruído. O segundo resultado é referente apenas à técnica MWF para redução de ruído, sem nenhuma técnica de preservação da lateralidade baseada na ITD da fonte de ruído. Para ambos azimutes a mediana é arrastada em direção ao azimute do sinal de fala (0°). A técnica 1 no azimute de $+60^\circ$ ficou com a mediana em torno de $+20^\circ$, no mesmo hemisfério onde a fonte de ruído estava localizada. Entretanto, apresentou alguns resultados no hemisfério contrário da posição original da fonte de ruído. No experimento com a fonte de ruído no azimute de -60° o resultado da técnica 1 foi ainda pior, a mediana dos resultados ficou em um azimute positivo, demonstrando que os voluntários consistentemente inverteram o hemisfério original.

A técnica 2a apresentou um surpreendente resultado para o cenário S_0N_{60} , com a mediana mais próxima do azimute real do que a técnica 2b. Em contrapartida, no experimento com o cenário S_0N_{-60} apresentou mediana mais próxima do azimute da fala (0°) do que a própria técnica MWF. No experimento com o cenário S_0N_{-60} a técnica 2b apresentou os melhores resultados entre todas as técnicas.

Os resultados apresentados pela técnica 3 mostraram que para ambos azimutes a mediana se manteve relativamente próxima do azimute real da fonte de ruído. Por fim, a técnica definida segundo os resultados objetivos como a que obteria os melhores resultados de lateralização (M.L.O) confirmou a expectativa e apresentou resultados com a menor variabilidade e mediana muito próximas do azimute original da fonte de ruído original. O fato de comprometer consideravelmente métricas importantes de qualidade, conforto acústico e inteligibilidade prejudica o uso dessa técnica em aplicações reais. Entretanto, os resultados da técnica 3 (M.L.O.) são importantes para mostrar o potencial da técnica proposta no presente trabalho para a preservação da lateralização de fontes sonoras pontuais.

5.2.1.2.2 Análise estatística dos resultados

Para avaliar a significância dos testes subjetivos, uma análise estatística foi realizada nos resultados mostrados na Figura 50 e na Figura 51. O método utilizado é o *Mann-Whitney-Wilcoxon* para análise de variância. Esse teste é não paramétrico para duas populações com amostras independentes e foi aplicado com nível de significância de 5%. A técnica testa a hipótese nula H_0 se os dados das duas técnicas são amostras de distribuições contínuas com medianas iguais. O teste assume que as duas amostras são independentes. Se o resultado for

$h = 0$, indica falha para rejeitar a hipótese nula enquanto que $h = 1$ indica rejeição da hipótese nula.

Tabela 8 – Resumo da análise estatística das técnicas avaliadas com experimentos subjetivos.

Comparação entre técnicas	Azimute = -60°	Azimute = $+60^\circ$
	h	h
N.P. × MWF	1	1
N.P. × Técnica 1	1	1
N.P. × Técnica 2a	0	0
N.P. × Técnica 2b	0	0
N.P. × Técnica 3	0	0
N.P. × Técnica 3 (M.L.O)	0	0

De acordo com os resultados mostrados na Tabela 8, os experimentos em ambos azimutes indicaram que o MWF e a técnica 1 não possuem a mesma mediana que os áudios não processados. Já as medianas das técnicas 2a, 2b e 3 podem ser consideradas idênticas.

5.2.2 Experimento 2

No experimento 1, foi verificado que as técnicas 2a e 2b (J_{MSC}) apresentaram resultados inferiores aos da técnica 3 (J_{IC}) nos testes subjetivos. Esse fato, aliado com a necessidade de ajustar dois parâmetros (β e γ), que dificulta sobremaneira sua utilização, fez com que ela não fosse avaliada no experimento 2. Por isso, apenas o desempenho da técnica MWF-IC, agora renomeada como J_C foi medido e comparado com as técnicas MWF-ITD renomeada como J_T , MWF renomeada como J_W , e sinais não processados (Raw) através de um conjunto de métricas objetivas e experimentos psicoacústicos para uma fonte de ruído pontual. Para ampliar a análise das técnicas, dois novos azimutes de -30° e $+30^\circ$ foram incluídos, gerando duas novas cenas acústicas de S_0N_{-30} e S_0N_{30} ainda em campo livre. Devido à similaridade dos resultados objetivos que apresentaram as técnicas nos cenários S_0N_{-30} e S_0N_{30} e nos cenários S_0N_{-60} e S_0N_{60} , apenas os resultados S_0N_{30} e S_0N_{60} serão apresentados. Além disso, o número de voluntários para os testes subjetivos foi ampliado de 5 para 15.

Tabela 9 – Resumo das soluções exploradas no experimento 2.

Técnica	Função Custo	Equação	Azimute	Leg.
1 (MWF-ITD)	J_T	(52)	+30°	----
			+60°	—
3 (MWF-IC)	J_C	(100)	+30°	----
			+60°	—

As fontes acústicas apresentam elevação nula, correspondendo ao plano transversal à cabeça do voluntário. A fonte de fala foi situada no azimute $\theta_S = 0^\circ$, enquanto que a fonte de ruído foi posicionada em 4 azimutes diferentes $\theta_N = \{-60^\circ, -30^\circ, 30^\circ, 60^\circ\}$ em frente ao voluntário.

O SNR e MOS-LQO da fala contaminada (*Raw*) para os quatro cenários acústicos estudados são apresentados na Tabela 10.

Tabela 10 – Medidas objetivas para o sinal de fala contaminada.

	S_0N_{-30}	S_0N_{-60}	S_0N_{30}	S_0N_{60}
MOS-LQO _L	1,1	1,1	1,2	1,2
MOS-LQO _R	1,2	1,2	1,1	1,1
SNR _L [dB]	0	0	4,6	4,5
SNR _R [dB]	4,6	4,5	0	0

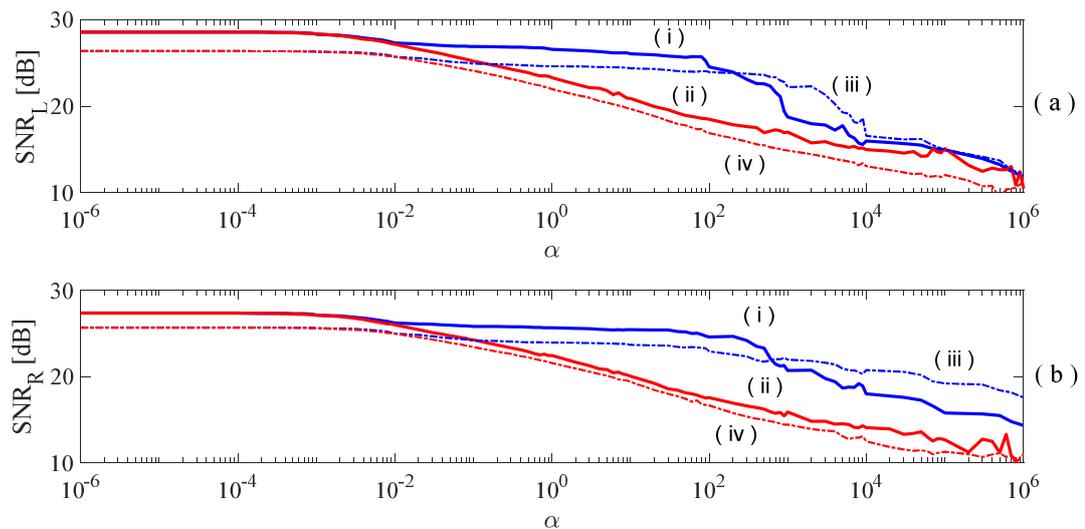
5.2.2.1 Avaliação objetiva

Cinco medidas objetivas foram calculadas para avaliar o desempenho das técnicas MWF-IC e MWF-ITD: a relação sinal ruído (SNR), que, quando comparada à SNR do sinal contaminado, mede a redução de ruído; o ganho de inteligibilidade na relação sinal ruído (Δ iSNR) (CHISAKI *et al.*, 2007), que estima a inteligibilidade do sinal de fala; o MOS-LQO (ITU, 2000), que estima a qualidade da fala contaminada processada (HU e LOIZOU, 2008), o erro da diferença de tempo interaural (Δ ITD) (BOGAERT *et al.*, 2007), calculado até 1,5 kHz, que mede a preservação da ITD; e o erro de coerência quadrática média (Δ MSC) (MARQUARDT, HOHMANN e DOCLO, 2013), que mede a variação da coerência entre os sinais de entrada e saída. Subíndices foram adicionados referindo-se à fala (S) e ruído (N) e esquerda (L) e direita (R).

5.2.2.1.1 SNR (Conforto acústico)

A Figura 52 apresenta a SNR nas orelhas esquerda e direita como função do fator de ponderação renomeado de (β) para (α_C) para MWF-IC e (α_T) para MWF-ITD para os cenários S_0N_{30} and S_0N_{60} em campo livre. Claramente, em ambas as técnicas a SNR diminui com o aumento do parâmetro de ponderação. A curva do MWF-IC começa a diminuir para α_C menores quando comparado com a MWF-ITD para os mesmos valores de α_T , que poderia levar a desqualificar prematuramente a técnica. O platô da curva localizado do lado esquerdo na figura corresponde ao SNR provido pela técnica MWF ($\alpha_C \rightarrow 0$ e $\alpha_T \rightarrow 0$).

Figura 52 – SNR na orelha esquerda (a) e direita (b) para técnica 1 (azul) e técnica 3 (vermelho) e S_0N_{60} (linha contínua) e S_0N_{30} (linha tracejada) nos cenários (i) J_T para S_0N_{60} ; (ii) J_C para S_0N_{60} ; (iii) J_T para S_0N_{30} ; (iv) J_C para S_0N_{30} .



5.2.2.1.2 Inteligibilidade e Qualidade

O mesmo comportamento da SNR foi observado para ambos Δ ISNR e MOS-LQO, respectivamente apresentadas na Figura 53 e Figura 54. A queda mais abrupta das curvas dessas métricas por parte da técnica MWF-IC é maior à medida que α_C aumenta, quando comparada com a técnica MWF-ITD.

Figura 53 – ΔISNR nas orelhas esquerda (a) e direita (b) para MWF-ITD (azul) e MWF-IC (vermelho). Cenários S_0N_{60} (linha contínua) e S_0N_{30} (linha tracejada): (i) J_T para S_0N_{60} ; (ii) J_C para S_0N_{60} ; (iii) J_T para S_0N_{30} ; (iv) J_C para S_0N_{30} .

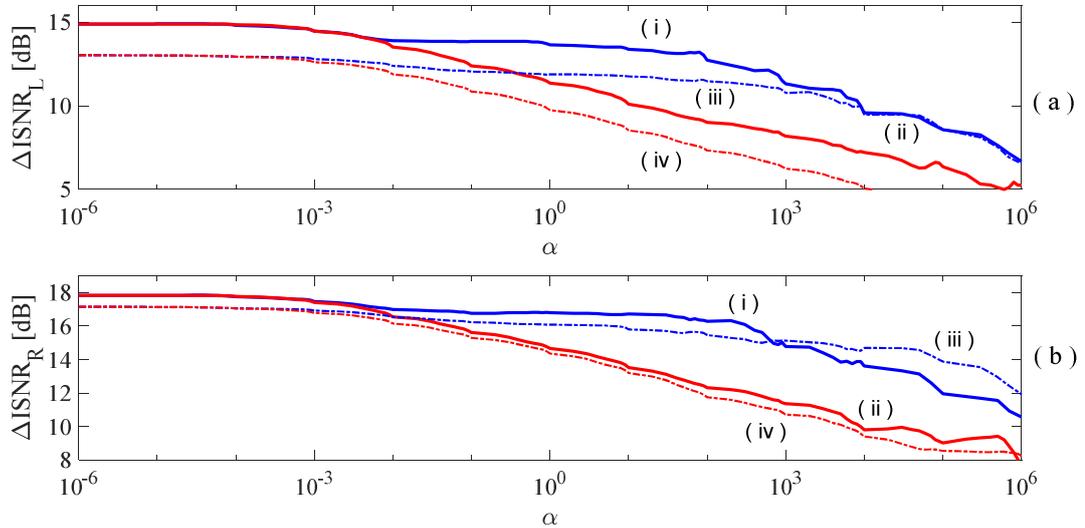
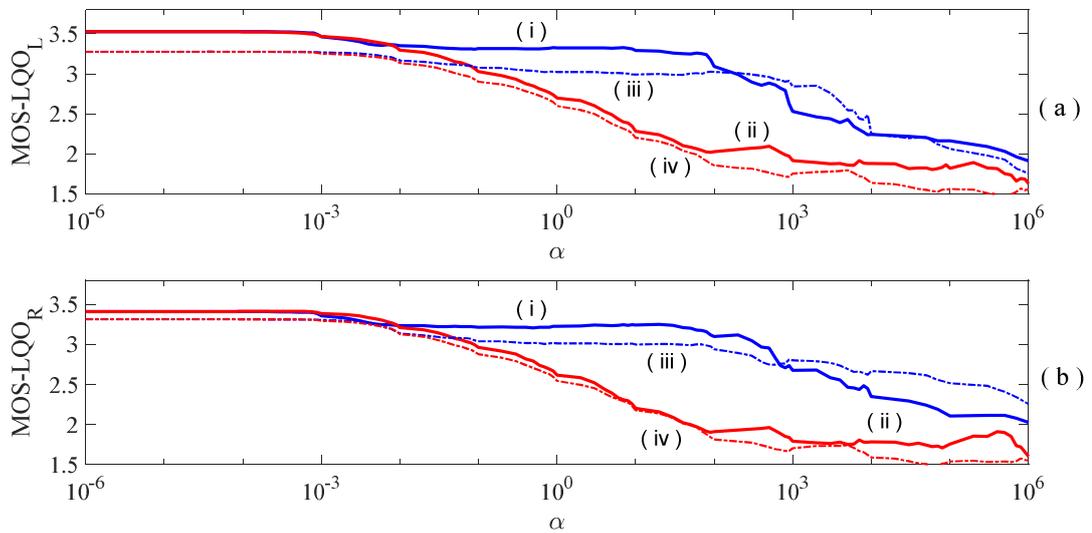


Figura 54 – MOS-LQO nas orelhas esquerda (a) e direita (b) para as técnicas MWF-ITD (azul) e MWF-IC (vermelho) nos cenários S_0N_{60} (linha contínua) e S_0N_{30} (linha tracejada): (i) J_T para S_0N_{60} ; (ii) J_C para S_0N_{60} ; (iii) J_T para S_0N_{30} ; (iv) J_C para S_0N_{30} .



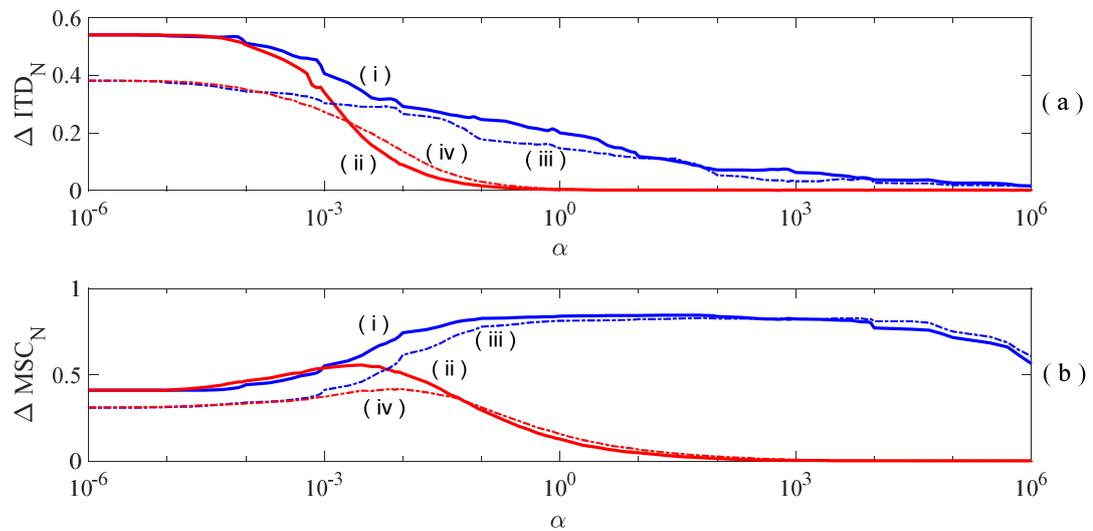
5.2.2.1.3 Erro de ITD (ΔITD) e Erro de MSC (ΔMSC) da fonte de ruído

A Figura 55(a) apresenta ΔITD , que indica que, quando comparado ao MWF convencional (J_W), ambos MWF-IC e MWF-ITD são capazes de reduzir a variação da diferença

de tempo interaural entre os sinais de entrada e saída à medida que os parâmetros de ponderação aumentam.

A Figura 55(b) mostra que o aumento do parâmetro de ponderação da MWF-ITD não reduz de maneira consistente a ΔMSC . Por outro lado, o MWF-IC reduz essa mesma métrica de maneira significativa restaurando a coerência interaural original do ruído residual.

Figura 55 – ΔITD_N do ruído (a) e ΔMSC_N do ruído para técnica 1 (azul) e técnica 3 (vermelho) e S_0N_{60} (linha contínua) e S_0N_{30} (linha tracejada) nos cenários (i) J_T para S_0N_{60} ; (ii) J_C para S_0N_{60} ; (iii) J_T para S_0N_{30} ; (iv) J_C para S_0N_{30} .



As medidas objetivas SNR, $\Delta ISNR$, MOS-LQO e ΔITD mostradas na Figura 52, Figura 53, Figura 54 e Figura 55 mostram claramente uma relação de compromisso entre redução de ruído e preservação espacial como uma função do fator de ponderação. Isso significa que conforto acústico, inteligibilidade e qualidade são degradados enquanto a ITD do ruído é preservada, com o aumento de α , em ambas as técnicas MWF-ITD e MWF-IC. Entretanto, o ΔMSC , mostrado na Figura 55 (b), indica que as técnicas MWF e MWF-ITD podem mudar substancialmente a coerência interaural dos sinais processados, independente da escolha do valor de α_T . Por outro lado, o MWF-IC reduz consistentemente tanto o ITD quanto o ΔMSC , quando α_C aumenta.

A Tabela 11 e Tabela 12 apresentam as medidas objetivas tanto para MWF-ITD quanto para MWF-IC, considerando os parâmetros de ponderação projetados para uma perda máxima de redução de ruído de 15% na pior orelha quando comparados ao MWF. O SNR_R foi reduzido em 13% (MWF-ITD) e 15% (MWF-IC) para o cenário S_0N_{30} ($\alpha_T = 3 \times 10^3$, $\alpha_C = 0,8$) e 11%

(MWF-ITD) e 13% (MWF-IC) para S_0N_{60} ($\alpha_T = 400$, $\alpha_C = 0,4$). A diferença entre as SNRs resultantes para ambas técnicas, MWF-ITD e MWF-IC, não excedeu o chamado *just-noticeable difference* de 3 dB (MCSHEFFERTY, WHITMER e AKEROYD, 2006). Da mesma forma, ambas as técnicas MWF-ITD e MWF-IC apresentaram aproximadamente a mesma qualidade na pior orelha para ambos cenários, uma vez que diferenças menores do que 0,2 de MOS-LQO não são perceptíveis pelos voluntários (SERVETTI e MARTIN, 2005). As pistas binaurais da fala, são apresentadas com o subíndice S, foram mantidas sem distorção (ΔITD_S e $\Delta MSC_S < 0,1$), enquanto que as pistas binaurais do ruído são apresentadas com o subíndice N, ΔITD_N foi reduzida consideravelmente de 0,4 para menos do que 0,1. A diferença mais significativa entre MWF-ITD e MWF-IC foi com relação à ΔMSC_N . A ΔMSC_N do ruído processado pelo MWF, que originalmente foi calculada como 0,4, sofreu um aumento para 0,8 com o MWF-ITD enquanto que o MWF-IC reduziu para 0,1 no cenário de S_0N_{30} e para 0,2 no cenário de S_0N_{60} . Isso indica que, diferentemente do MWF-ITD, o método MWF-IC controla tanto as pistas binaurais de ITD quanto de IC.

Tabela 11 – Medidas objetivas para S_0N_{30} : $\alpha_T = 3 \times 10^3 (J_T)$, $\alpha_C = 0,8 (J_C)$.

	J_W	J_T	J_C	
S_0N_{30}	SNR _L [dB]	26,3	21,3	22
	SNR _R [dB]	24,7	21,6	20,9
	$\Delta ISNR_L$ [dB]	13,4	10,6	10,3
	$\Delta ISNR_R$ [dB]	16,6	15,0	14,3
	MOS-LQO _L	3,4	2,8	2,7
	MOS-LQO _R	2,9	2,8	2,4
	ΔITD_S	2×10^{-2}	5×10^{-2}	2×10^{-2}
	ΔITD_N	0,4	3×10^{-2}	4×10^{-3}
	ΔMSC_S	7×10^{-3}	5×10^{-3}	2×10^{-2}
	ΔMSC_N	0,4	0,82	0,1

Tabela 12 – Medidas objetivas para S_0N_{60} : $\alpha_T = 400$ (J_T), $\alpha_C = 0,4$ (J_C).

	J_W	J_T	J_C	
S_0N_{60}	SNR _L [dB]	28.5	22.7	24.2
	SNR _R [dB]	26.3	23.5	22.9
	Δ ISNR _L [dB]	14.7	12.1	12.0
	Δ ISNR _R [dB]	17.8	15.7	15.6
	MOS-LQO _L	3.6	2.9	3
	MOS-LQO _R	3.1	3	2.7
	Δ ITD _S	2×10^{-2}	5×10^{-2}	3×10^{-2}
	Δ ITD _N	0.43	7×10^{-2}	1×10^{-2}
	Δ MSC _S	1×10^{-2}	1×10^{-2}	2×10^{-2}
	Δ MSC _N	0.43	0.82	0.2

5.2.2.2 Avaliação psicoacústica

Na avaliação subjetiva do experimento 2 foi utilizado o mesmo laptop e fone de ouvido do experimento 1. Além disso, foram selecionados 15 voluntários (11 homens e 4 mulheres) com idades entre 19 e 39 anos (média de 29 anos e desvio padrão de 4,9 anos). Nenhum voluntário relatou algum problema auditivo prévio. O procedimento aplicado nessa avaliação subjetiva é o mesmo aplicado no experimento 1 e composto pelas fases de a) aprendizado, b) treinamento e c) teste. Na primeira fase, os voluntários escutaram apenas o sinal de ruído filtrado por HRIRs relativas a sete azimutes diferentes $\{-90^\circ, -60^\circ, -30^\circ, 0^\circ, 30^\circ, 60^\circ, 90^\circ\}$. Cada áudio foi sincronizado com uma informação visual (mesma interface gráfica do experimento 1) relacionada com o azimute real do ruído processado. Na fase de treinamento, os voluntários foram desafiados a identificar o azimute dos mesmo sete áudios, apresentados em ordem aleatória, sem qualquer conhecimento prévio sobre o real azimute. Nessa fase, voluntários que cometeram inversões de hemisfério em suas avaliações (erros de lateralização relacionados aos lados direito e esquerdo) foram excluídos do experimento. Na fase de testes os voluntários classificaram um conjunto de 16 áudios apresentados de forma aleatória. Um transferidor virtual com 13 caixas, numa escala entre -90° e 90° em degraus de 15° , foi apresentado na tela do *laptop*. Os áudios selecionados compreendem 4 ruídos apresentados na fase de treinamento ($-60^\circ, -30^\circ, 30^\circ, 60^\circ$), assim como as versões filtradas do ruído presente

nos sinais de fala contaminada, para os cenários S_0N_{-30} , S_0N_{30} , S_0N_{-60} e S_0N_{60} de acordo com os coeficientes ótimos obtidos com J_W , J_T e J_C .

Visando a proporcionar uma comparação justa de desempenho entre as técnicas, uma redução de ruído de 15% (máximo), quando comparada ao MWF foi deliberadamente aplicada tanto para as técnicas MWF-IC e MWF-ITD na “pior orelha”. Os parâmetros de ponderação alocados em cada função custo J_C e J_T , respeitando essa condição, são apresentados na Tabela 13.

Tabela 13 – Fator de ponderação para uma perda de redução de ruído de no máximo 15% comparado com a técnica MWF.

	$\alpha_T(J_T)$	$\alpha_C(J_C)$
S_0N_{-30}/S_0N_{30}	3×10^3	0,8
S_0N_{-60}/S_0N_{60}	4×10^2	0,4

Na Figura 56 os espectrogramas dos sinais analisados na melhor orelha dentro do cenário S_0N_{60} são apresentados. Podemos verificar que o mesmo mostra uma considerável redução de ruído tanto pelo MWF, MWF-ITD e MWF-IC sem notáveis diferenças visuais entre elas.

A Figura 57 mostra os resultados da fase de treinamento de 15 voluntários na forma de diagramas de caixa (*boxplot*). Pode-se observar que nenhum voluntário cometeu erros de inversão de hemisfério durante essa fase. A caixa é representada pelo retângulo azul que limita os quartis superior e inferior denominados de q_1 e q_3 (25% e 75%), respectivamente, enquanto a linha vermelha dentro da caixa indica a mediana. Além disso, a caixa representa a amplitude interquartil (IIQ), uma medida estatística de dispersão entre q_3 e q_1 ($IIQ = q_3 - q_1$). Os *whiskers* (linhas retas pretas e tracejadas que se estendem verticalmente a partir da caixa) indicam os maiores e menores valores das amostras. Os valores discrepantes (*outliers*) são representados por cruces vermelhas correspondendo às amostras maiores do que $q_3 + \varpi(q_3 - q_1)$ e menores do que $q_1 - \varpi(q_3 - q_1)$. A variável ϖ é definida pelo valor padrão de 1,5 (SHANNON, 2003) e representa os extremos superiores e inferiores, os quais não são considerados como valores discrepantes.

Figura 56 – Espectrograma dos sinais na orelha direita para o cenário S_0N_{60} : (a) fala limpa; (b) fala contaminada; (c) fala contaminada processada pelo MWF; (d) fala contaminada processada pelo MWF-ITD; (e) fala contaminada processada pelo MWF-IC.

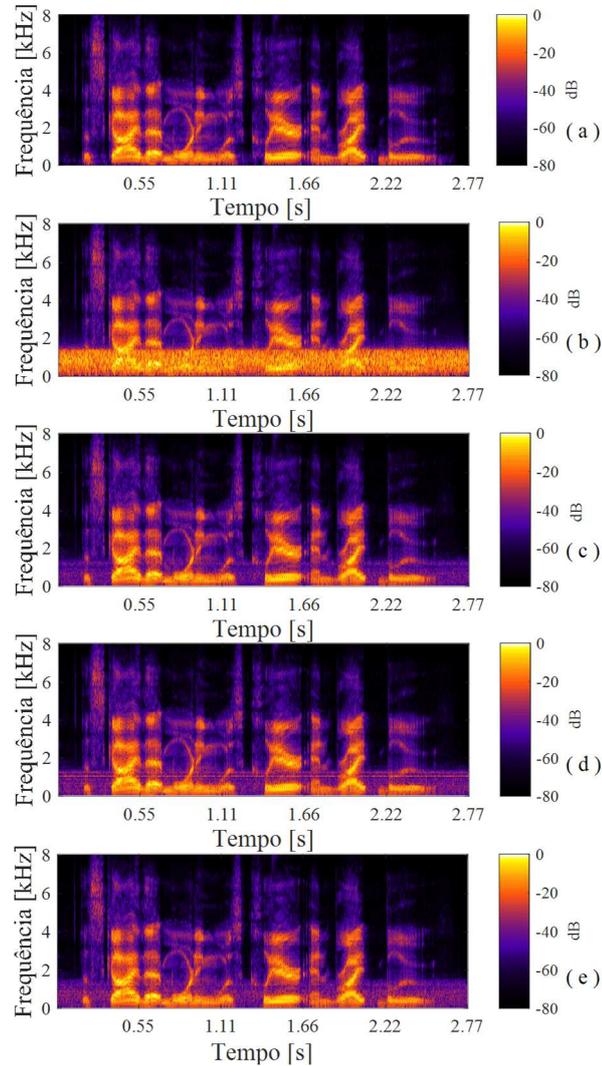
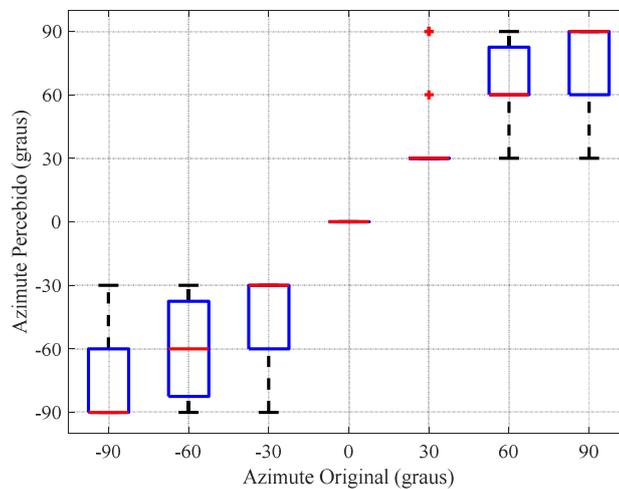


Figura 57 – Classificação dos resultados obtidos na fase de treinamento para os 15 voluntários com o ruído com frequência abaixo de 1,5 kHz.



A Figura 57 atesta que os 15 voluntários apresentaram uma capacidade de julgamento de lateralização adequada. A mediana amostral calculada concorda precisamente com o azimuth real da fonte sonora. Além disso, nenhuma inversão de hemisfério foi apresentada.

5.2.2.2.1 Lateralização da fonte de ruído

A Figura 58 e Figura 59 mostram os resultados dos 15 voluntários selecionados na fase de teste para os cenários S_0N_{-30} , S_0N_{30} , S_0N_{-60} e S_0N_{60} . Diagramas de caixa modificados foram utilizados para facilitar a interpretação dos resultados. Nessas figuras, a localização dos voluntários é representada pelo diagrama de uma cabeça humana (vista superior) usando um fone de ouvido. A fonte de fala é representada por uma outra cabeça menor em frente ao voluntário com uma distância de d_s metros. A fonte de ruído é indicada através de um símbolo específico contendo o azimuth real (negrito). Os diagramas de caixa são representados por caixas preenchidas com cores diferentes de acordo com uma legenda predeterminedada.

Figura 58 – Boxplot modificado para o experimento psicoacústico 2. Percepção do azimuth da fonte de ruído. (a) S_0N_{30} (cima), e (b) S_0N_{60} (baixo): Ruído processado por MWF-IC (azul), MWF-ITD (vermelho), MWF (verde), e ruído não processado (amarelo).

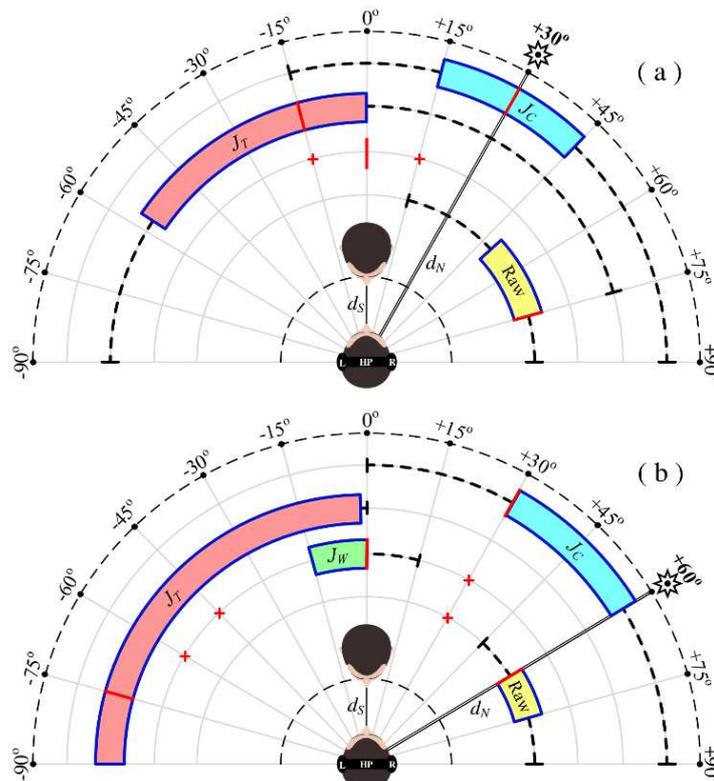
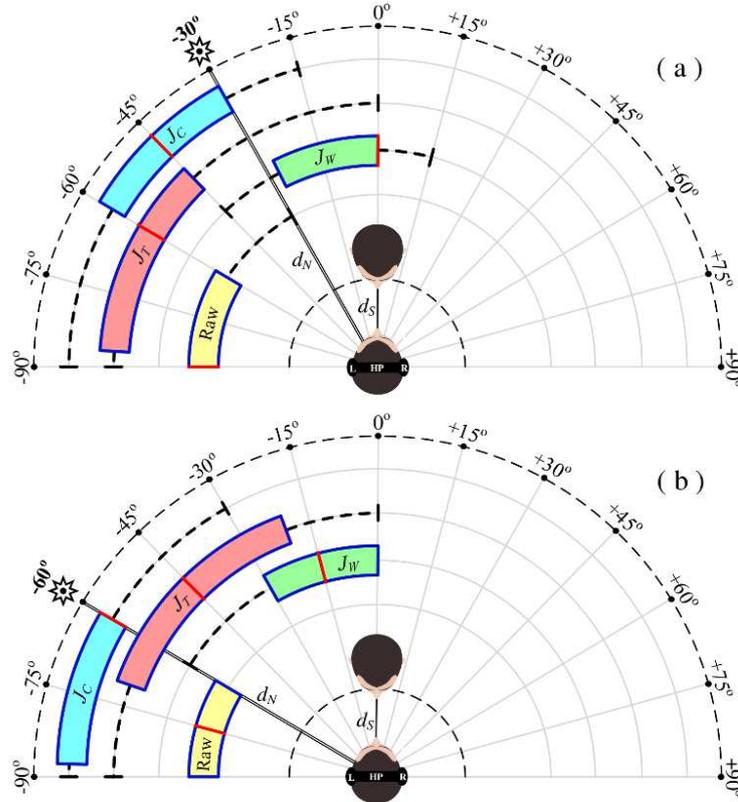


Figura 59 – Boxplot modificado para o experimento psicoacústico 2. Percepção do azimuth da fonte de ruído. (a) S_0N_{-30} (cima), e (b) S_0N_{-60} (baixo): Ruído processado por MWF-IC (azul), MWF-ITD (vermelho), MWF (verde), e ruído não processado (amarelo).



Como esperado, o MWF (caixa verde) apresentou uma polarização em seus resultados para o azimuth do sinal de fala (0°) (CORNELIS *et al.*, 2010). Também é verificado que o azimuth médio do ruído não processado (caixa amarela) é levado para valores de azimuth maiores do que os reais. Isso pode ser explicado pela concordância natural entre as pistas binaurais de ITD e ILD, que podem amplificar a percepção de lateralização. Esse fenômeno não ocorre com os ruídos processados, uma vez que as técnicas não controlam a pista de ILD.

A Tabela 14 apresenta a média amostral, desvio padrão e mediana para os diagramas de caixa mostrados na Figura 58 e Figura 59. A Tabela 15 apresenta as diferenças absolutas entre a média e mediana para o MWF-ITD e MWF-IC quando comparadas com o ruído original em todos os cenários.

Tabela 14 – Experimento psicoacústico: Média amostral do azimute (\bar{x}_i), desvio padrão (s_i) e mediana (\tilde{x}_i) tal que $i = \{ R, W, T, C \}$, significando respectivamente: não processado (RAW), e processado por MWF, MWF-ITD e MWF-IC.

		S_0N_{-30}	S_0N_{30}	S_0N_{-60}	S_0N_{60}
RAW	\bar{x}_R	-74°	65°	-75°	62°
	s_R	21°	21°	$12,7^\circ$	$14,9^\circ$
	\tilde{x}_R	-90°	75°	-75°	60°
J_W	\bar{x}_W	-11°	1°	-18°	-8°
	s_W	$19,2^\circ$	9°	19°	$27,7^\circ$
	\tilde{x}_W	0°	0°	-15°	0°
J_T	\bar{x}_T	-59°	-27°	-44°	-53°
	s_T	$28,6^\circ$	44°	30°	$41,2^\circ$
	\tilde{x}_T	-60°	-15°	-45°	-75°
J_C	\bar{x}_C	-48°	33°	-67°	41°
	s_C	$22,1^\circ$	$27,1^\circ$	$19,5^\circ$	$22,3^\circ$
	\tilde{x}_C	-45°	30°	-60°	30°

Tabela 15 – Experimento psicoacústico: Média entre as diferenças absolutas (\bar{x}_i) e mediana (\tilde{x}_i) dos azimutes para MWF-ITD (T) e MWF-IC (C) quando comparados com o azimute percebido do ruído RAW (R).

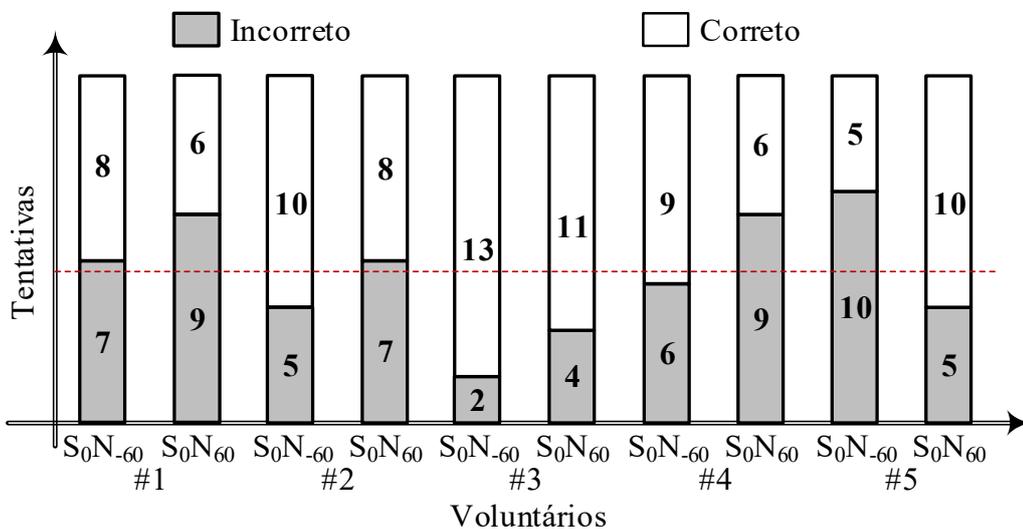
		S_0N_{-30}	S_0N_{30}	S_0N_{-60}	S_0N_{60}	Média
J_T	$ \bar{x}_R - \bar{x}_T $	15°	92°	31°	115°	$63,3^\circ$
	$ \tilde{x}_R - \tilde{x}_T $	30°	90°	30°	135°	$71,3^\circ$
J_C	$ \bar{x}_R - \bar{x}_C $	26°	32°	8°	21°	$21,8^\circ$
	$ \tilde{x}_R - \tilde{x}_C $	45°	45°	15°	30°	$33,8^\circ$

Os resultados mais impressionantes foram obtidos com a técnica MWF-ITD tanto no cenário S_0N_{30} quanto no S_0N_{60} . De acordo com a Figura 58, praticamente todos os voluntários cometeram inversões de hemisfério em suas avaliações e conseqüentemente, a média amostral e mediana apontaram para o hemisfério oposto da real localização da fonte de ruído.

Visando a esclarecer esses resultados, um novo experimento subjetivo foi realizado com os 5 voluntários que obtiveram o melhor resultado na etapa de treinamento. Primeiramente, eles escutaram 15 realizações diferentes do ruído processado pelas técnicas MWF-ITD e MWF-

IC nos cenários S_0N_{60} e S_0N_{-60} . Os voluntários classificaram cada áudio de acordo com o hemisfério percebido (Esquerda ou Direita). Novamente os voluntários não cometeram inversões de hemisfério com ruído processado pela técnica MWF-IC. Além disso, todos os voluntários apresentaram inversões de hemisférios em seus julgamentos nos áudios processados pela técnica MWF-ITD. A Figura 60 retrata os resultados obtidos para a técnica MWF-ITD. A linha vermelha tracejada indica o número médio de erros de lateralização (6,4) para as 15 avaliações.

Figura 60 – Inversões de lateralização obtidas nos experimentos psicoacústicos para técnica MWF-ITD: 5 voluntários e 15 tentativas para cada.

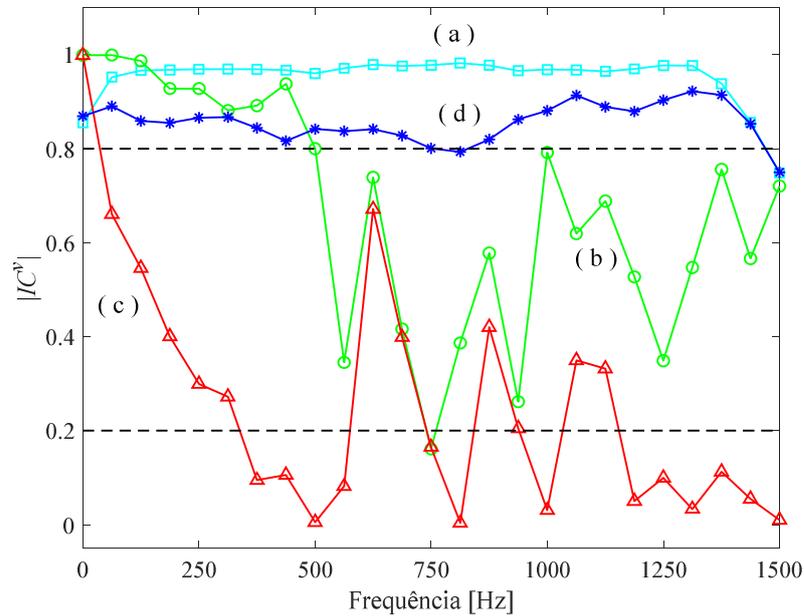


Duas importantes observações podem ser enfatizadas: primeiro o MWF-IC apresentou estimativas mais exatas (mediana) e precisas (IQR) dos azimutes reais quando comparado com MWF-ITD (Tabela 14), para todos os cenários acústicos. Em segundo lugar, enquanto o MWF-IC sempre fornece a localização de hemisfério correta, o MWF-ITD resulta em um significativo número de inversões de hemisfério, isso é corroborado pela Figura 60.

A importância da IC para uma efetiva preservação da pista de ITD foi previamente analisada em RAKERD e HARTMANN (2010). Esse trabalho avaliou a capacidade dos voluntários em lateralizar fontes sonoras em ambientes reverberantes para diferentes valores de $|IC_{in}^v|$. Ruídos acústicos na banda efetiva da ITD ($< 1,5$ kHz) foram investigados. Aproximadamente 50% de inversões de hemisfério foram observadas para cenários com $|IC_{in}^v| < 0,2$, enquanto que não houve inversões de hemisfério para $|IC_{in}^v| > 0,8$. Esses resultados mostram a relação entre a magnitude da coerência interaural e a habilidade que os seres

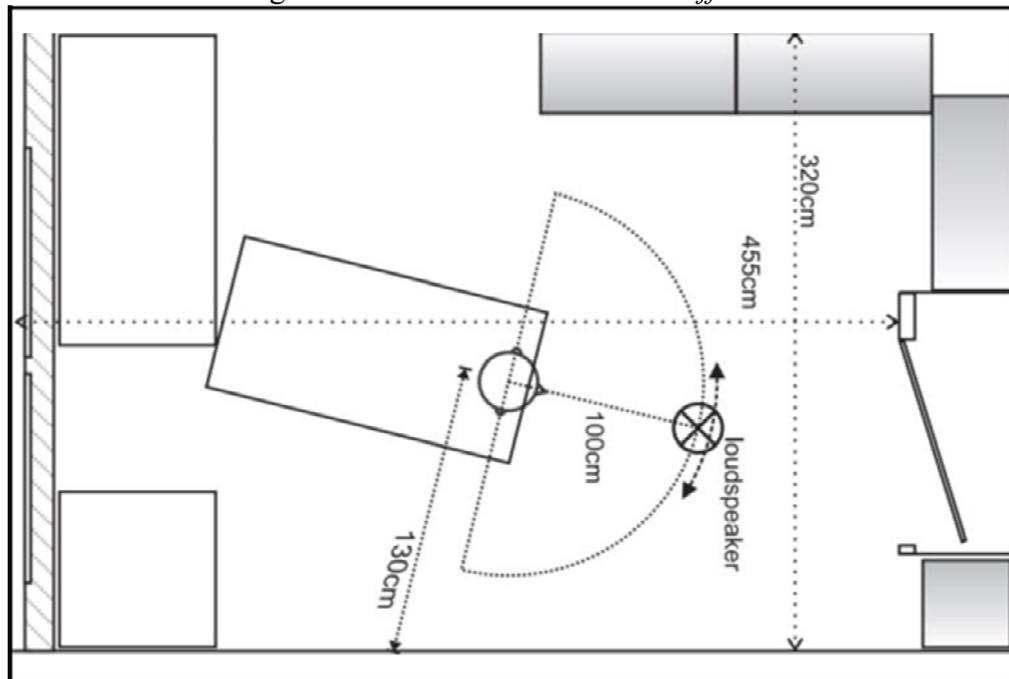
humanos possuem de lateralizar sons em campo livre usando a informação de ITD. Um fenômeno equivalente pode ser observado nos resultados do MWF-IC apresentados nessa seção. A Figura 61 mostra um exemplo da magnitude da coerência interaural como função da frequência para o sinal de ruído aplicado em nossos experimentos. A elevada magnitude da coerência associada ao ruído não processado $|IC^v_{in}(k)|$ (linha amarela - \square) é característica de um cenário anecoico. O ruído processado pelo MWF-IC (linha azul - $*$) preservou a magnitude original da coerência interaural, basicamente resultando em $|IC^v_{out}(k)| > 0,8$. Entretanto, ambas as técnicas MWF e MWF-ITD resultaram em uma redução na magnitude da coerência. De fato, o MWF-ITD resultou valores muito pequenos, alcançando $|IC^v_{out}(k)| < 0,2$ para altas frequências. Essa escala de magnitude da coerência é uma característica de campos sonoros difusos, encontrados em ambientes com alta reverberação. Do ponto de vista teórico, a magnitude da coerência do ruído na saída deveria ser unitária, devido às características de processamento linear do vetor dos coeficientes $\mathbf{w}_L(k)$ and $\mathbf{w}_R(k)$. A redução de $|IC^v_{out}|$ observada é devida provavelmente aos erros de estimação da STFT, resultando em $\Phi_{\mathbf{v}\mathbf{v}}(k) = \mathbf{h}_v(k)\mathbf{h}_v^H(k)\sigma_v^2(k) + \Delta$, no qual $\mathbf{h}_v(k)$ são as HRIRs relacionadas com a localização da fonte de ruído e Δ é uma matriz de erro (hermitiana) de posto completo. Ainda na Figura 61, as linhas escuras tracejadas mostram os valores limites para a magnitude da coerência interaural descrita em RAKERD e HARTMANN (2010). Em HARTMANN, RAKERD e CRAWFORD (2016), experimentos psicoacústicos de lateralização indicaram que pequenas magnitudes de IC aumentam a importância da ILD até mesmo em frequências abaixo de 1,5 kHz. Frequências com informação dissonante de ITD e ILD podem contribuir para erros de lateralização. O alto número de inversões de hemisfério verificado na MWF-ITD pode ser atribuído a uma combinação de uma pequena coerência interaural com pistas binaurais de ILD e ITD discordantes.

Figura 61 – Magnitude da coerência interaural: (a) ruído não processado (quadrado amarelo); (b) ruído processado pelo MWF (círculo verde); (c) ruído processado pelo MWF-ITD (triângulo vermelho); (d) ruído processado pelo MWF-IC (asterisco azul).



5.3 CENÁRIO ACÚSTICO COM REVERBERAÇÃO

Embora cenários acústicos sem reverberação sejam comuns em situações cotidianas, existem também situações (ambientes fechados) onde as reflexões das ondas sonoras (reverberação) reduzem a capacidade de localizar/lateralizar a correta posição de fontes sonoras. Para demonstrar a capacidade da técnica proposta nessa situação, um experimento subjetivo foi conduzido nessas condições para os cenários S_0N_{60} e S_0N_{-60} com reverberação (tempo de reverberação $RT_{60} = 300$ ms). O cenário escolhido é conhecido como *Office 1* e apresenta as dimensões mostradas na Figura 62 com ambas as fontes sonoras a uma distância de 1 metro do indivíduo e elevação de 0° . As HRTFs utilizadas foram obtidas em KAYSER *et al.*, (2009).

Figura 62 – Cenário reverberante *Office I*.

Um grupo de 10 voluntários, sem nenhuma queixa prévia de problemas de audição, participou do experimento cujos resultados são mostrados na Figura 63 para a configuração S_0N_{-60} e para a configuração S_0N_{60} na Figura 64 em forma de *boxplots*. O ruído processado pela técnica J_C é mostrado em azul, MWF-ITD mostrado em vermelho, MWF em verde e o ruído não processado em amarelo. O SNR na pior orelha é de 0dB. Esses experimentos mostraram que, de fato, a reverberação reduz o desempenho do MWF-IC (J_C) em comparação ao cenário anecoico. Contudo, o MWF-IC ainda é capaz de preservar o cenário acústico quando comparado com a técnica MWF-ITD (J_T) que mesmo em ambientes reverberantes continua apresentando inversões de hemisfério.

Figura 63 – Experimento subjetivo com 10 voluntários para o cenário S_0N_{-60} em condições a) anecoicas e b) reverberantes.

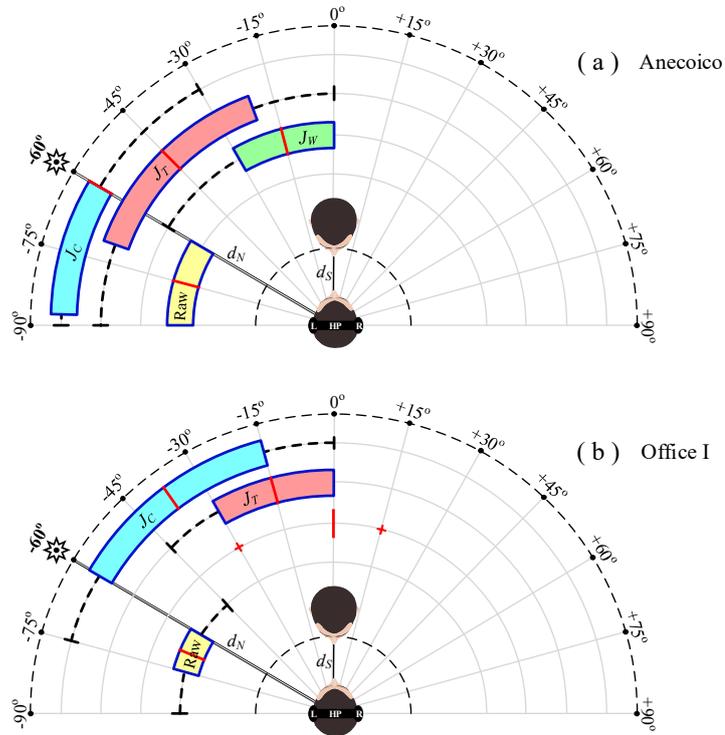
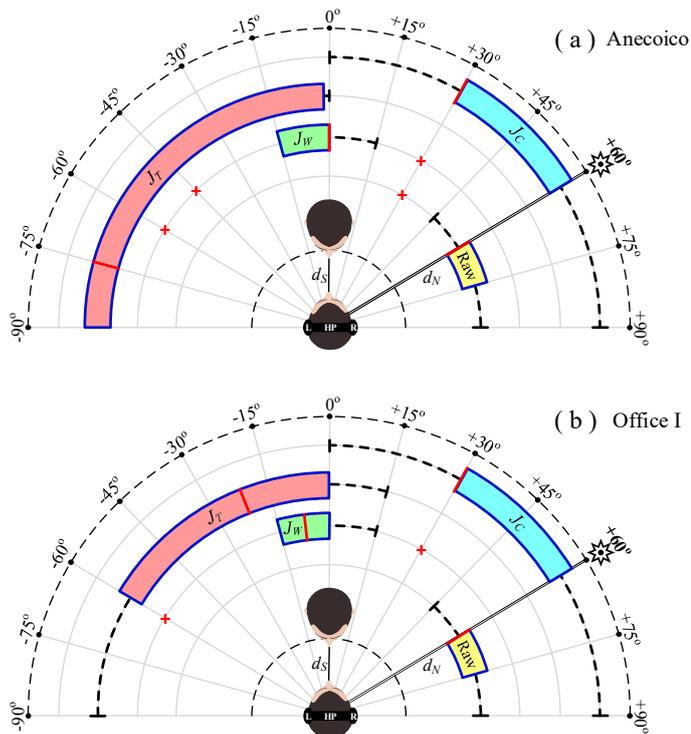


Figura 64 – Experimento subjetivo com 10 voluntários para o cenário S_0N_{60} em condições a) anecoicas e b) reverberantes.



Os resultados obtidos indicam que o MWF-IC é uma técnica promissora para redução de ruído com preservação espacial tanto para fontes de ruído difusas (MARQUARDT e DOCCLO, 2018) quanto para pontuais, como demonstrado nesse trabalho.

Finalmente, a utilização simultânea das funções custo da IC e ILD pode levar a uma técnica de redução de ruído com a preservação perceptivamente relevante do cenário espacial sob a escala completa de frequências auditivas. Esse é um tópico que será investigado em trabalhos futuros.

6 CONCLUSÃO

O presente trabalho propõe o uso da coerência interaural como um método para obter uma preservação perceptivamente relevante da diferença de tempo interaural em aparelhos auditivos binaurais. Foi mostrado que a técnica MWF-ITD distorce a coerência natural de fontes de ruído pontual obtidas por microfones em ambientes do tipo campo livre e reverberante. O método MWF-IC fornece uma relação de compromisso entre redução de ruído e preservação da espacialização original do ruído para frequências abaixo de 1,5 kHz. Isso ocorre devido à preservação simultânea das pistas de ITD e IC. Medidas objetivas e experimentos psicoacústicos corroboraram a análise teórica, indicando que a preservação da IC é fundamental para a percepção subjetiva da ITD e correta lateralização de fontes sonoras direcionais. Os resultados obtidos são de interesse especial aos projetistas de aparelhos auditivos na busca por técnicas binaurais de redução de ruído que preservem o cenário acústico original tanto para fontes de ruído difusas quanto pontuais em frequências abaixo de 1,5 kHz.

7 PRINCIPAIS CONTRIBUIÇÕES DO TRABALHO

As principais contribuições inovadoras do presente trabalho são:

- Fornecer evidências experimentais consistentes e suporte teórico de que a técnica MWF-ITD, originalmente proposta em DOCLO *et al.*, (2005b), não é capaz de preservar a lateralização original de fontes sonoras;
- Mostrar que a minimização da diferença entre a IC de entrada e saída de um sinal produzido por uma fonte acústica pontual corresponde à minimização da diferença entre a ITD de entrada e saída;
- Propor a técnica MWF-IC, originalmente derivada para campos acústicos difusos, como um método eficiente para redução de ruído e preservação da ITD de fontes sonoras pontuais;
- Apresentar resultados objetivos e experimentos psicoacústicos que corroboram os fundamentos anteriores, mostrando que a técnica de redução de ruído MWF-IC resulta em uma preservação perceptualmente relevante das pistas de localização baseadas na ITD.

8 TRABALHOS FUTUROS

Os resultados e conclusões apresentados nesse trabalho abrem uma série de possibilidades de trabalhos envolvendo a temática apresentada. Dentre elas podemos elencar:

- Comparação da técnica MWF-IC+ILD com a técnica MWF-ITF: Foi mostrado neste trabalho que a técnica MWF-IC é capaz de preservar a lateralização de fontes de ruído pontuais até 1,5kHz. Estudos prévios também mostraram que a técnica MWF-ILD é capaz de preservar a lateralização de ruído acima dessa frequência. Sendo assim, existe uma possibilidade de comparação de desempenho em termos de redução de ruído da técnica MWF-ITF com as técnicas MWF-IC+ILD para ruídos de banda larga;
- Implementação *on-line* da técnica MWF-IC: A técnica *on-line* do MWF-ILD já foi publicada em (CARMO e COSTA, 2018). Com a implementação *on-line* do MWF-IC seria possível expandir o trabalho em três frentes diferentes:
 1. Cenários móveis: Nesse trabalho utilizamos ambas fontes de fala e ruído estáticas em azimutes pré-estabelecidos. Entretanto, em situações cotidianas, é muito comum que as fontes sonoras (principalmente o ruído) se movimentem dentro do cenário acústico. Nesse caso uma implementação *on-line* permitiria recalcular regularmente as matrizes de coerência do ruído permitindo extrair características dinâmicas desses cenários. Dessa forma seria possível gerar novos coeficientes responsáveis por manter a lateralização de uma fonte de ruído pontual em movimento;
 2. Ampliação na gama de ruídos: Como já discutido anteriormente, todos experimentos desenvolvidos no trabalho contaram com fontes de ruído estacionário com frequência inferior a 1,5 kHz. A implementação *on-line* permitiria explorar ruídos com características estatísticas diferentes;
 3. Implementações em tempo real: com esse tipo de abordagem abrem-se aplicações voltadas para sistemas embarcados de tempo real. Esse tipo de implementação aproxima a técnica de aplicações reais voltadas para aparelhos auditivos binaurais.

- Aplicar restrições de ILD mesmo na banda abaixo de 1,5kHz visando melhorar a percepção subjetiva de lateralização: Experimentos mostraram que a ILD, desde que concordante com a ITD, é capaz de melhorar resultados em experimentos de lateralização na banda onde a ITD, segundo a teoria duplex, é a responsável pela lateralização de fontes sonora pontuais, conforme mostrado na Seção 2.2.5. Entretanto, abaixo de 1,5 kHz, a técnica MWF-IC não controla nenhum aspecto relacionado com a ILD. Por isso, nessas frequências é possível que ocorram soluções onde a ILD discorda da ITD original e degrada os resultados experimentais.

REFERÊNCIAS

- ACAR, B. et al. Effects of hearing aids on cognitive functions and depressive signs in elderly people. **Archives of Gerontology and Geriatrics (Elsevier)**, n. 52, p. 250-252, Jan. 2011.
- ACOUSTICAL SOCIETY OF AMERICA. **American National Standard Methods for Calculation of the speech intelligibility index**. [S.l.]. 1997.
- ARORA, J. S. **Introduction to Optimum Design**. 2nd. ed. [S.l.]: Elsevier, 2004.
- AUDIOMAX. Centro Auditivo Audiomax. **O que é uma audiometria?**, 2016. Disponível em: <<http://grupoaudiomax.com.br/v1/o-que-e-uma-audiometria/>>. Acesso em: 14 mar. 2017.
- BENESTY, J.; MAKINO, S.; CHEN, J. **Speech Enhancement**. [S.l.]: Springer, 2005.
- BERNSTEIN, L. R.; TRAHOTIS, C. Lateralization of low-frequency complex waveforms: the use of envelope-based temporal disparities. **The Journal of the Acoustical Society of America** , v. 77, p. 1868-1880, 1985.
- BERTEIN, J. C.; CESCHI, R. **Discrete Stochastic Processes and Optimal Filtering**. 1st. ed. [S.l.]: Wiley-ISTE, 2007.
- BLAUERT, J. (Ed.). **The technology of Binaural Listening**. [S.l.]: Springer, 2013.
- BLAUERT, J. **Spatial Hearing: The Psychophysics of Human Sound Localization**. Cambridge: The MIT Press, 1997.
- BOGAERT, T. V. et al. **Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter**. Proc. do ICASSP. Honolulu, HI, USA: [s.n.]. 2007. p. 565-568.
- BORISAGAR, K.; THANKI, R.; SEDANI, B. **Speech Enhancement Techniques for Digital Hearing Aids**. 1st. ed. [S.l.]: Springer International Publishing, 2019.
- CARLILE, S. **Virtual auditory space: Generation and applications**. [S.l.]: Springer Science & Business Media, 2013.
- CARMO, D. M. **Algoritmo Adaptativo de Redução de Ruído e Preservação das Diferenças de nível Interaural para Aparelhos Auditivos**. UFSC. [S.l.], p. 110p. 2017.
- CARMO, D. M.; COSTA, M. H. Online approximation of the multichannel Wiener filter with preservation of interaural level difference for binaural hearing-aids. **Computers in Biology and Medicine**, v. 95, p. 188-197, 2018.
- CHISAKI, Y. et al. Real-Time processing using the frequency domain binaural model. **Appl. Acoust.**, v. 68, p. 923-938, 2007.

- CORNELIS, B. et al. Theoretical Analysis of Binaural Multimicrophone Noise Reduction Techniques. **IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING**, v. 18, n. 2, p. 342-355, Fev. 2010.
- CORNELIS, B.; MOONEN, M.; WOUTERS, J. Performance analysis of multichannel Wiener filter-based noise reduction in hearing aids under second order statistics errors. **IEEE Trans. Audio Speech Lang. Process.**, v. 19, n. 5, p. 1368-1381, 2011.
- COSTA, M. H.; NAYLOR, P. A. ILD PRESERVATION IN THE MULTICHANNEL WIENER FILTER FOR BINAURAL HEARING AID APPLICATIONS. **European Signal Processing Conference (EUSIPCO)**, 2014. 636-640.
- CROCHIERE, R. E. A weighted overlap-add method of short-time Fourier Analysis/Synthesis. **IEEE Transactions on Acoustics, Speech, and Signal Processing**, n. 1, p. 99-102, Feb. 1980.
- DAVIS, H.; SILVERMAN, R. S. **Hearing and Deafness**. New York: Rinehart & Wiston, 1970.
- DILLON, H. **Hearing Aids**. 1st. ed. [S.l.]: Thieme, 2001.
- DILLON, H. **Hearing Aids**. 2nd. ed. [S.l.]: Thieme, 2012.
- DOCLO, S. et al. **Extension of the Multi-Channel Wiener Filter with localization cues for noise reduction in binaural hearing aids**. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). [S.l.]: [s.n.]. 2005a. p. 221-224.
- DOCLO, S. et al. **Extension of The Multi-Channel Wiener Filter with ITD cues for Noise Reduction in Binaural Hearing Aids**. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA). [S.l.]: [s.n.]. 2005b. p. 70-73.
- DOCLO, S. et al. Frequency-domain criterion for speech distortion weighted multichannel Wiener for Robust Noise reduction. **Speech Communication**, v. 49, n. 7-8, p. 636-656, Jul-Ago 2007.
- DOCLO, S. et al. Handbook on array processing and sensor networks. [S.l.]: [s.n.], 2010. p. 269 - 302.
- EHIMA. EHIMA - European Hearing Instrument Manufactures Association. Disponivel em: <www.ehima.com>. Acesso em: 20 Julho 2016.
- FAIRNIE, J.; MOORE, B. C.; REMINGTON, A. Missing a trick: Auditory load modulates conscious awareness in audition, v. 42, n. 7, p. 930, 2016.
- GANNOT, S.; BURSHTEIN, D.; WEINSTEIN, E. Signal enhancement using Beamforming and non-stationarity with applications to speech. **IEEE Trans. Signal Processing**, v. 49, n. 8, p. 1614-1626, Agosto 2001.

- GREENBERG, J. E.; PETERSON, P. M.; ZUREK, P. M. Intelligibility-weighted measures of speech-to-interface. **The Journal of the Acoustical society of America**, v. 94, p. 3009-3010, Nov. 1993.
- HABETS, E.; NAYLOR, P. A. **An online quasi-newton algorithm for blind SIMO identification**. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP). [S.l.]: [s.n.]. 2010. p. 2662-2665.
- HADAD, E.; DOCCLO, S.; GANNOT, S. The Binaural LCMV Beamformer and its Performance Analysis. **IEEE/ACM Transactions on Audio, Speech and Language Processing** , v. 24, n. 3, p. 543 - 558, Março 2016.
- HARTMANN, W. M.; MACAULAY, E. J. Anatomical limits in interaural time differences: an ecological perspective. **Frontiers in Neuroscience**, v. 8, n. 34, p. 1-13, Fev. 2014.
- HARTMANN, W. M.; RAKERD, B.; CRAWFORD, Z. D. Transaural experiments and a revised duplex theory for the localization of low-frequency tones. **J. Acoust. Soc. Am.**, v. 139, n. 2, p. 968-985, Fev. 2016.
- HAYKIN, S. et al. **Method and device for binaural signal enhancement**. WO2007028250A2, 13 maio 2007.
- HU, Y.; LOIZOU, P. C. Evaluation of objective quality measures for speech enhancement. **IEEE Trans. Audio Speech Lang. Process.** , v. 16, p. 229-238, 2008.
- HUANG, Y.; BENESTY, J. A class of frequency-domain adaptive approaches to blind multichannel identification. **IEEE Transactions on signal processing**, v. 51, n. 1, p. 11-24, 2003.
- IEEE Xplore Digital Library. Disponível em: <<http://ieeexplore.ieee.org/Xplore/home.jsp>>.
- ITU. **Perceptual evaluation of Speech quality (PESQ), and objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs**. ITU-T Recommendation P. 862. [S.l.]. 2000.
- ITU. **Mapping function for transforming P.862 raw result**. ITU-T Recommendation P.862.1. [S.l.]. 2003.
- ITU. **Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs**. ITU-T Recommendation P.862.2. [S.l.]. 2007.
- JEFFRESS, L. A. A place theory of sound localization. **Journal of comparative and Physiological Psychology**, n. 41, p. 35-39, Feb 1948.
- JENSEN, J.; PEDERSEN, M. **Analysis of beamformer directed single-channel noise reduction system for hearing aid applications**. ICASSP. [S.l.]: [s.n.]. 2015. p. 5728 - 5732.

- KAYSER, H. et al. Database of Multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. **Eurasip Journal on Advances in Signal Processing**, p. 1-10, 2009.
- KLASEN, T. J. et al. **Binaural Noise Reduction for Hearing Aids: Preserving Interaural Time Delay Cues**. IEEE BENELUX/DSP Valley Signal Processing Symposium (SPS DARTS). Antuérpia/BEL: [s.n.]. 2005. p. 23-26.
- KLASEN, T. J. et al. Binaural Multi-Channel Wiener Filtering for Hearing Aids: Preserving Interaural Time and Level Differences. **IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**, 2006. 145 - 148.
- KOUTROUVELIS, A. I. et al. Relaxed binaural LCMV beamforming. **IEEE/ACM Transactions on Audio, Speech and Language Processing**, v. 25, n. 1, p. 137-152, 2017.
- LIN, F. R.; FERRUCCI, L. Hearing Loss and Falls Among Older Adults in The United States. **Archives of Internal Medicine**, v. 172, n. 4, p. 369-371, Março 2012.
- LOIZOU, P. **Speech Enhancement: Theory and Practice**. 2. ed. [S.l.]: CRC Press, 2013.
- LUTS, H. et al. Multicenter evaluation of signal enhancement algorithms for hearing aids. **The Journal of the Acoustical Society of America**, v. 127, p. 1491-1505, 2010. ISSN 3.
- MACPHERSON, E. A.; MIDDLEBROOKS, J. C. Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. **The Journal of the Acoustics Society of America**, n. 111, 2002.
- MALTBY, M. **Occupational Audiometry: Monitoring and protecting hearing at work**. 1st. ed. [S.l.]: Elsevier, 2005.
- MARQUARDT, D. et al. Theoretical Analysis of Linearly Constrained Multi-Channel Wiener Filtering Algorithms for Combined Noise Reduction and Binaural Cue Preservation in Binaural Hearing Aids. **IEEE/ACM Transactions on Audio, Speech, and Language Processing**, v. 23, n. 12, p. 2384-2397, Dez 2015.
- MARQUARDT, D.; DOCLO, S. Interaural coherence preservation for binaural noise reduction using partial noise estimation and spectral postfiltering. **IEEE Trans. Audio Speech Lang. Process.**, v. 26, n. 7, p. 1261-1274, 2018.
- MARQUARDT, D.; HOHMANN, V.; DOCLO, S. **COHERENCE PRESERVATION IN MULTI-CHANNEL WIENER FILTERING BASED NOISE REDUCTION FOR BINAURAL HEARING AIDS**. ICASSP. [S.l.]: [s.n.]. 2013. p. 8648-8652.
- MCSHEFFERTY, D.; WHITMER, W. M.; AKEROYD, M. A. The just-meaningful difference in speech-to-noise ratio. **Trends Hear.**, v. 20, p. 1-11, 2006.

- MED-EL. How does hearing works., 2012. Disponível em: <<http://www.medel.com/us/show/index/id/60/title/How-HearingWorks?PHPSESSID=d2slqgcle7d9tr606vidtt3m65>>. Acesso em: 28 Mar 2015.
- MERRIAN-WEBSTER.COM. **Merriam-webster**, 2017. Disponível em: <<https://www.merriam-webster.com>>. Acesso em: 8 Maio 2017.
- MILLS, A. W. Lateralization of High frequency tones. **The Journal of the Acoustics Society of America**, n. 32, p. 132-134, 1960.
- PHONAK. Phonak Dalia, 2012. Disponível em: <https://www.phonak.com/content/phonak/us/en/support/product-support/hearing-aids/further-hearing-aids/_jcr_content/par/downloads_15/productdownloadspar/download_1/file.res/029-0077-03_PhonakDalia_custom.pdf>. Acesso em: 18 Abril 2015.
- PHONAK. Phonak TAO Q, 2013. Disponível em: <https://www.phonak.com/content/dam/phonak/HQ/en/solution/products/tao_q/documents/user_guide_tao_q.pdf>. Acesso em: 18 Abril 2015.
- PLENGE, G. On the differences between localization and lateralization. **The Journal of the Acoustical Society of America**, v. 56, n. 3, p. 944-951, Setembro 1974.
- POPELKA, G. R. et al. **Hearing Aids**. 1st. ed. [S.l.]: Springer, 2016.
- RAKERD, B.; HARTMANN, W. Localization of Sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise. **The Journal of The Acoustical Society of America**, n. 128, Nov. 2010.
- RASPAUD, M.; VISTE, H.; EVANGELISTA, G. Binaural Source Localization by Joint Estimation of ILD and ITD. **IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING**, v. 18, n. 1, p. 68-77, Jan. 2010.
- RAYLEIGH, L. On our Perception of sound of sound direction. **Philos. Mag.**, v. 13, p. 214-232, Abril 1907.
- RIX, A. et al. **Perceptual evaluation of speech quality (PESQ) - A new method for speech quality assessment of telephone networks and codecs**. Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing. [S.l.]: [s.n.]. 2001. p. 749-752.
- ROGIN, C. M. **Markettrak 9: Exciting New Beginnings**. 64th Annual IHS convention & expo. Orlando, FL.: [s.n.]. 2015.
- SANDEL, T. T. et al. Localization of sound from single and paired sources. **The Journal of the Acoustical Society of America**, n. 27, p. 842-852, 1955.

- SAYERS, B. M. Acoustic image lateralization judgements with binaural tones. **The Journal of the Acoustical Society of America**, n. 36, p. 923-926, 1964.
- SERVETTI, A.; MARTIN, J. C. **Error tolerant MAC extension for speech communications over 802.11 WLANs**. Veh. Technol. Conf. (VTC). [S.l.]: [s.n.]. 2005. p. 1-5.
- SHANNON, D. Box-and-whisker plots with the SAS. **Pharm. Stat.**, v. 2, p. 291-295, 2003.
- SMITH, R. C. G.; PRICE, S. R. Modelling of Human Low Frequency Sound Localization Acuity Demonstrates Dominance of Spatial Variation of Interaural Time Difference and Sugests Uniform Just-Noticeable Differences in Interaural Time Difference. **Plos one**, v. 9, n. 2, Fevereiro 2014.
- SPRIET, A.; MOONEN, M.; WOUTERS, J. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction. **Signal Processing**, v. 84, n. 12, p. 2367-2387, 2004.
- STAAB, W. Hearing Health & Technology Matters. **Localization: More Important Than Word Recognition?**, 2015. Disponivel em: <<http://hearinghealthmatters.org/waynesworld/2015/localization-more-important-than-word-recognition/>>. Acesso em: 6 Outubro 2016.
- STEVENS, S. S.; NEWMAN, E. B. The location of actual sources of sound. **The Journal of the Acoustical Society of America**, n. 48, p. 297-306, 1936.
- STRUTT, J. W. On our perception of the direction of sound. **Proc. R. Soc. London**, v. 83, p. 61-64, 1909.
- TATE, M. **Principles of Hearing Aid Audiology**. 1st. ed. [S.l.]: Springer-Science+Business Media, 1994.
- TEXAS EAR CENTER. Disorders of the ear. **Texas Ear Center**, 2017. Disponivel em: <<http://www.texasearcenter.com/disorders-of-the-ear>>. Acesso em: 13 junho 2015.
- TRAHIOTIS, C.; BERNSTEIN, L. R.; AKEROYD, M. A. Manipulating the "straightness" and "curvature" of patterns of interaural cross correlation affects listeners sensitivity to changes in interaural delay. **The Journal of the Acoustical Society of America**, n. 109, p. 321-330, 2001.
- VEEN, B. V.; BUCKLEY, K. Beamforming: A versatile approach to spatial filtering. **IEEE ASSP Mag.**, v. 5, n. 2, p. 4-24, 1988.
- WANG, H. et al. Robust Adaptive Beamforming Based on Worst-Case and Norm Constraint. **International Journal of Antennas and Propagation**, 2015.
- WEB of science, 2017. Disponivel em: <<https://login.webofknowledge.com/>>.

WEIGMANN, E. Binaural Hearing and Bilateral Hearing Aid Care. **Siemens Audiologische Technik GmbH**, 2014. Disponível em:

<https://global.hearing.siemens.com/news/media/2014/01/Backgrounder_binaural-hearing_201209_en.pdf>. Acesso em: 6 Fevereiro 2015.

WIENER, N. **Extrapolation, Interpolation, and Smoothing of Stationary Time Series**. New York: John Wiley & Sons, 1949.

WIGHTMAN, F. L.; KISTLER, D. J. The dominant role of low-frequency interaural time differences in sound localization. **The Journal of the Acoustical Society of America**, n. 91, p. 1648-1661, 1992.

YOST, W. A. Lateral position of sinusoids presented with interaural intensive and temporal differences. **J. Acoust. Soc. Am.**, p. 397-409, 1981.

APÊNDICE A – Esquema utilizado nos experimentos psicoacústicos

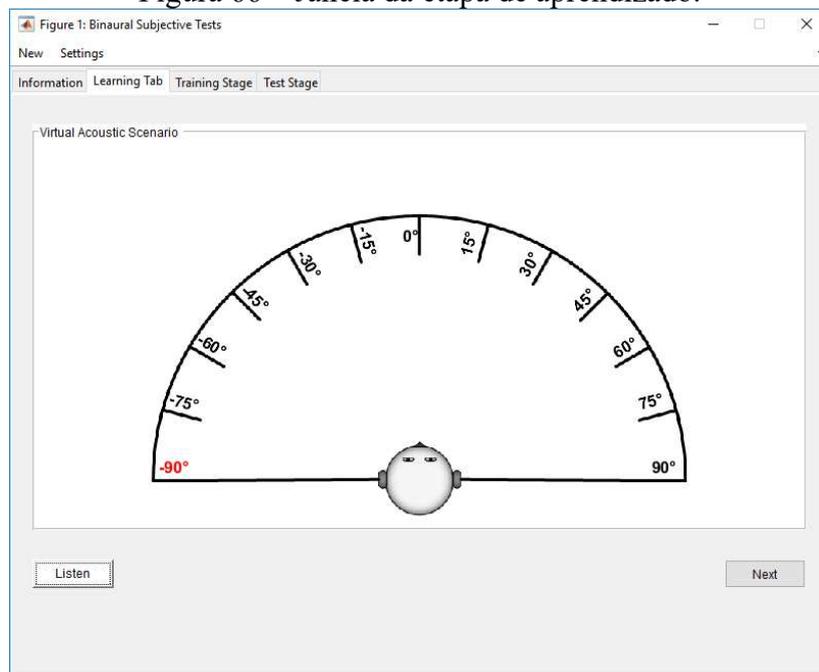
Os experimentos psicoacústicos realizados no trabalho foram desenvolvidos conforme o diagrama apresentado na Figura 65 e proposto em (CARMO, 2017).

Figura 65 – Diagrama do esquema aplicado nos experimentos psicoacústicos.



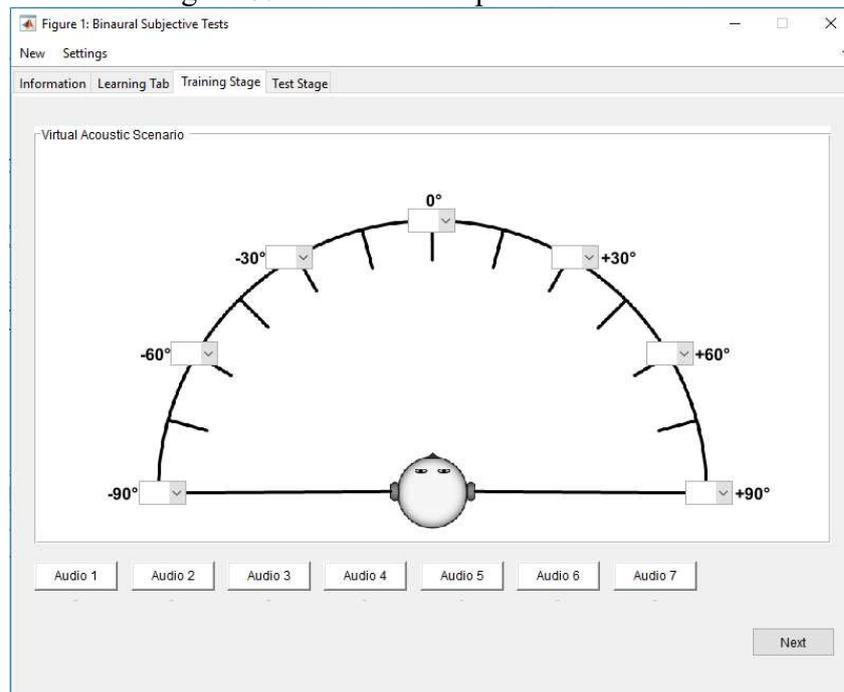
Inicialmente os voluntários foram submetidos a uma etapa de aprendizado, que consiste em escutar apenas o ruído nos 13 azimutes mostrados na Figura 66. Cada vez que o voluntário pressiona o botão “Listen”, um áudio de aproximadamente dois segundos é reproduzido no azimute destacado em vermelho. Dessa forma o voluntário associa o áudio do ruído binaural com os azimutes. Nessa etapa o voluntário apenas pressiona o botão e escuta a sequência completa (-90° até 90°). O voluntário pode repetir a sequência quantas vezes achar necessário até se sentir apto para passar para a etapa de treinamento.

Figura 66 – Janela da etapa de aprendizado.



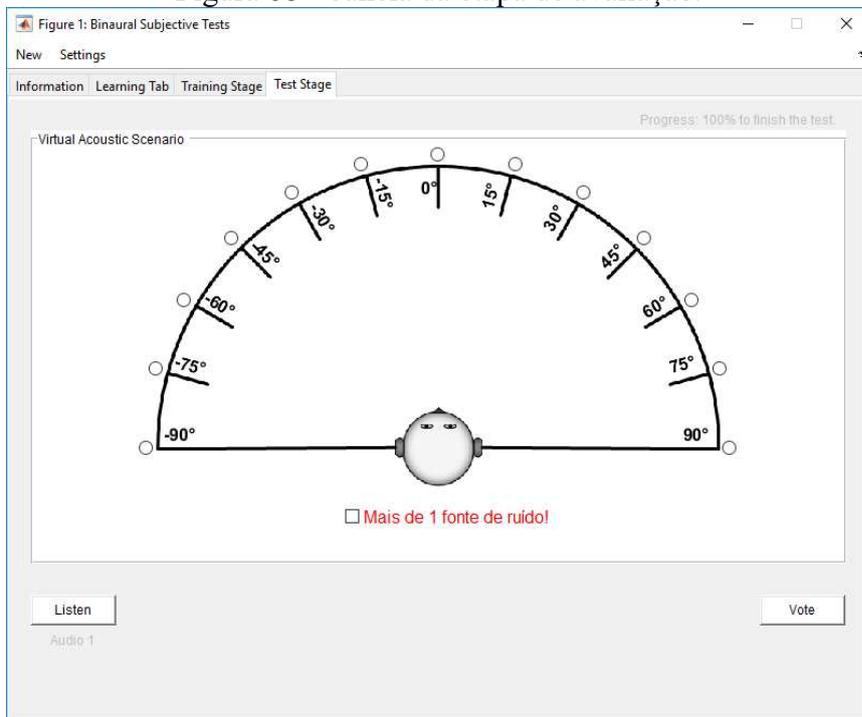
Na etapa de treinamento, o voluntário é desafiado a ordenar apenas sete dos trezes arquivos de áudio em seus respectivos azimutes. É permitido ouvir os arquivos quantas vezes for necessário, assim como trocar de azimute se achar necessário. Essa etapa é eliminatória e decide se os resultados do voluntário serão ou não aproveitados. O fator determinante para selecionar um voluntário está relacionado com a capacidade de lateralizar corretamente a fonte de ruído. A Figura 67, mostra a janela que o voluntário usa para ligar cada um dos sete áudios aos respectivos sete azimutes. O gabarito não é informado ao voluntário, e independente de seu resultado, o voluntário passa para a fase avaliação. O mesmo é considerado reprovado se inverter o azimute em ao menos um dos sete julgamentos. Em caso de reprovação, os resultados são simplesmente descartados. O motivo desse procedimento é evitar constrangimento ao voluntário.

Figura 67 – Janela da etapa de treinamento.



A última etapa consta na avaliação das técnicas exploradas no trabalho. A Figura 68 apresenta a janela na qual os voluntários votam no azimute desejado. Nessa etapa o voluntário pode registrar sua resposta com o auxílio do *mouse* para mover o cursor (em azul) com precisão de 5°.

Figura 68 – Janela da etapa de avaliação.



Todos os testes subjetivos executados no presente trabalho foram realizados de acordo com o projeto submetido ao comitê de ética da universidade sob responsabilidade do professor orientador do trabalho.

APÊNDICE B – Artigo publicado na IEEE/ACM TASLP

IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 27, NO. 4, APRIL 2019

753

Perceptually Relevant Preservation of Interaural Time Differences in Binaural Hearing Aids

Fábio Pires Itturriet [✉] and Márcio Holsbach Costa [✉], *Member, IEEE*

Abstract—This paper presents a noise reduction method with perceptually relevant preservation of the interaural time difference (ITD) of the residual noise in binaural hearing aids. The interaural coherence (IC) concept, previously applied to the multichannel Wiener filter (MWF) for preservation of the spatial subjective sensation of diffuse noise fields, is proposed here to both preserve and emphasize the ITD binaural cues of a directional acoustic noise source. It is demonstrated that the previously developed MWF-ITD technique may decrease the original IC magnitude of the processed noise, consequently increasing the variance of the interaural phase difference (IPD) of the output signals. It is shown that the MWF-IC technique simultaneously minimizes a nonlinear function of the difference between input and output IPD, which is strictly related to ITD, and preserves the natural coherence of directional noise captured by the reference microphones. Objective measures and psychoacoustic experiments corroborate the theoretical findings, showing that the MWF-IC technique provides relevant noise reduction, while preserving the original ITD subjective perception and original lateralization for a directional noise source. These results are especially relevant for hearing aid designers, since they indicate MWF-IC as a noise reduction technique that provides residual noise spatial preservation for both diffuse and directional noise sources in frequencies below 1.5 kHz.

Index Terms—Hearing aids, noise reduction, binaural, speech processing, Wiener filter.

I. INTRODUCTION

NOISE reduction algorithms are an important part of modern hearing aids. One of the major complaints of hearing aid users is poor speech intelligibility due to background noise. Many studies have demonstrated that hearing-impaired people need a SNR-50¹ from 10 to 30 dB higher than that required for the non-impaired [1]. This happens due to the loss of spectral resolution of the damaged auditory system [2]. Consequently, it may result in social isolation, professional difficulties and risk to personal safety. According to [3], more than 80 percent of the hearing impaired have both ears affected by a reduction in

hearing ability, requiring the simultaneous use of two hearing aids.

Bilateral hearing aids (left and right gadgets working independently) do not preserve the original acoustic localization cues, distorting the listener's sense of auditory space, as well as its ability to localize, separate, and track sound sources [4]. Although noise-reduction could be effective, localization of residual sounds is generally best achieved by turning-off the processing routines [2], [4], thus diminishing the equipment acceptability. This represents a major disadvantage to the hearing-aid user since the immediate localization of sources of interest is paramount to allow visual identification (traffic, safety warnings) and/or lip-reading.

Despite many advances in hearing assistive technology, noise reduction strategies that preserve spatial localization of sound sources remain challenging, mainly due to the difficulty of integrating different localization cues into the noise reduction framework. In this context, early approaches applied the same processing (attenuation factor) to both ears [5]–[7]. Nowadays, common approaches for noise reduction are the linearly constrained minimum variance beamformer and the generalized side-lobe canceller. However, these techniques rely on prior knowledge about source localization and/or head related transfer functions, presenting significant performance degradation when the assumed conditions deviate from real ones [8].

Binaural² Multichannel Wiener Filter (MWF) based techniques have been extensively explored in the current scientific literature [9]–[13]. This approach permits deep theoretical insights about its design and performance [14]–[16]. Although it was theoretically demonstrated that the conventional binaural MWF naturally preserves speech localization cues, its major drawback resides in residual noise at the output inheriting the input speech localization cues [14]. As a result, the hearing aid user cannot make use of the psychoacoustic mechanisms related to spatial separation between noise and speech sources [17] to mask unwanted information (better ear advantage) and therefore improve speech understanding or localization or tracking of noise sources [9], [18].

In order to overcome these issues, some MWF variations have been developed. They can be divided into two classes: in the first, controlled amounts of unprocessed signal are allowed at the output of the hearing aids. Although it was demonstrated that they result in better noise source spatial localization as compared to the conventional MWF [18], [19], this paradigm

Manuscript received July 16, 2018; revised November 4, 2018 and January 18, 2019; accepted January 23, 2019. Date of publication January 29, 2019; date of current version February 15, 2019. This work was supported by CAPES and CNPq under Grant 304867/2015-2. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Tan Lee. (Corresponding author: Márcio Holsbach Costa.)

F. P. Itturriet is with the Instituto Federal de Educação, Ciência e Tecnologia Sul-Rio-Grandense, Charqueadas 96745-000, Brazil (e-mail: fabio.itturriet@gmail.com).

M. H. Costa is with the Departamento de Engenharia Elétrica e Eletrônica, Universidade Federal de Santa Catarina, Florianópolis 88040-900, Brazil (e-mail: costa@eel.ufsc.br).

Digital Object Identifier 10.1109/TASLP.2019.2895973

¹SNR-50 is the signal to noise ratio needed for the comprehension of 50% of the speech in a conversation.

²When both (left and right) hearing aids exchange data.

is not strictly related to preservation of the localization cues. The second approach is characterized by adding extra terms to the MWF cost function to penalize solutions that do not preserve the desired binaural cues. It has been demonstrated that interaural time differences (ITD) (differences between transmission times in both ears) are the primary spatial cues in mammals and birds [20], followed by interaural level differences (ILD) (difference between magnitudes at each ear). Other cues are also accepted as playing secondary roles (providing supplementary information) in sound source localization, such as monaural spectral features provided by pinnae and non-acoustical cues such as source familiarity and vision [21]. For frequency components above 1500 Hz, ITD may exhibit ambiguity due to short wavelengths compared to the distance between the ears, and due to the breakdown of phase locking in the auditory neurons [22]–[24]. ILD is mainly determined by the attenuation provided by the head and torso (head shadow effect) and is more pronounced when the head is in between the source and one of the ears. Due to the acoustic properties (reflection, diffraction, scattering, interference and resonance) of the head, torso and pinnae, ILD presents a strong increasing dependence with frequency [22], [25]. When available, low-frequency ITD information is dominant over ILD and spectral shape information, which are used to resolve ambiguities [21].

The first attempt to preserve ITD and ILD of residual noise by means of inserting an auxiliary cost function into the MWF formulation was presented in [10]. The efficacy of MWF-ILD methods [10], [26] for preservation of the acoustic scenario was supported by early amplitude stereo panning techniques that have demonstrated that ILD carries enough information for creating complex artificial auditory scenes even in headphones [27]. Differently from the MWF-ILD, and despite the apparent physical appropriateness of the proposed estimator and associated cost function, extensive experiments performed by the authors of this work indicated that the MWF-ITD [28] does not provide perceptually relevant results in preserving the localization of the residual noise. This fact can be only partially explained by the observations presented in [29]. As a result, this is still an important open research area.

In this work, we propose the use of the interaural coherence (IC) measure for preserving and emphasizing the ITD localization cues in MWF-processed signals. The IC was firstly proposed in [30] for preserving the original spatial characteristics of diffuse noise fields, and in [31] it was shown that IC is a nonlinear function of ITD. Here, it is shown that by minimizing IC differences between input and processed signals, the original ITD localization cues are also preserved, resulting in the correct psychoacoustic impression of the original acoustic scene for directional sources. The novel contributions of this work are as follows: 1) it provides strong experimental evidence and theoretical support that the MWF-ITD technique, as originally described in [10], is not capable of preserving the original lateralization of the processed noise; 2) it shows that minimizing the difference between input and output IC of a signal produced by a directional acoustic source corresponds to minimizing the difference between input and output ITD; 3) we propose that the MWF-IC technique, originally derived for diffuse fields, is an

efficient noise reduction method for providing ITD preservation of directional sound sources; 4) finally, this study provides objective results and psychoacoustic experiments that corroborate the previous rationales, showing that the MWF-IC noise reduction technique leads to perceptually relevant preservation of the ITD localization cues.

The remainder of this paper is structured as follows: the binaural problem formulation is presented in Section II, while Section III introduces the MWF noise reduction technique. A brief review of binaural cost functions for preserving the binaural cues is presented in Section IV. In Section V, the MWF-IC technique is proposed as an efficient noise reduction technique with ITD preservation of directional sources. The experimental setup is described in Section VI, and the results are presented in Section VII. Finally, sections VIII and IX present the discussion and conclusions of this work.

Throughout this document, bold uppercase and lowercase letters represent matrices and vectors, respectively, while italics represent scalars.

II. SIGNAL AND SYSTEM MODELS

The application context considered in this paper comprises a binaural fitting of hearing aids, working in full-duplex mode without bit-rate limitations. The operating scenario assumes the existence of one acoustic source of interest $x(t)$ (speech) and one interfering noise source $v(t)$. Both sources are assumed as having a fixed (or slowly varying) position in a given time-window. Frequency domain decomposition is applied to the incoming signals through an N -bin Short-Time Fourier Transform (STFT). For a sampling frequency of f_s samples per second, for each time-frame λ and frequency k , for the M_L left microphones and the M_R right microphones, the received signals are defined as follows:

$$\begin{aligned} y_{L,l}(\lambda, k) &= x_{L,l}(\lambda, k) + v_{L,l}(\lambda, k) \\ y_{R,r}(\lambda, k) &= x_{R,r}(\lambda, k) + v_{R,r}(\lambda, k), \end{aligned} \quad (1)$$

where x is the speech signal, v is the noise, with $l = 1, \dots, M_L$, and $r = 1, \dots, M_R$. The collection of these received signals can be expressed in vector form as follows

$$\mathbf{y}(\lambda, k) = \mathbf{x}(\lambda, k) + \mathbf{v}(\lambda, k), \quad (2)$$

in which $\mathbf{y}(\lambda, k) = [y_{L,1}(\lambda, k) \dots y_{L,M_L}(\lambda, k) y_{R,1}(\lambda, k) \dots y_{R,M_R}(\lambda, k)]^T$, $\mathbf{x}(\lambda, k) = [x_{L,1}(\lambda, k) \dots x_{L,M_L}(\lambda, k) x_{R,1}(\lambda, k) \dots x_{R,M_R}(\lambda, k)]^T$ and $\mathbf{v}(\lambda, k) = [v_{L,1}(\lambda, k) \dots v_{L,M_L}(\lambda, k) v_{R,1}(\lambda, k) \dots v_{R,M_R}(\lambda, k)]^T$ are vectors with dimension $M \times 1$ with $M = M_L + M_R$.

Considering the deterministic vectors \mathbf{q}_L and \mathbf{q}_R , both with dimensions $M \times 1$, which contain 1 in the element corresponding to the respective (left/right) reference microphone and zeros otherwise, the reference vectors of the hearing aids (without processing) are given by the following:

$$\begin{aligned} y_{L,ref}(\lambda, k) &= x_{L,ref}(\lambda, k) + v_{L,ref}(\lambda, k) = \mathbf{q}_L^T \mathbf{y}(\lambda, k) \\ y_{R,ref}(\lambda, k) &= x_{R,ref}(\lambda, k) + v_{R,ref}(\lambda, k) = \mathbf{q}_R^T \mathbf{y}(\lambda, k). \end{aligned} \quad (3)$$

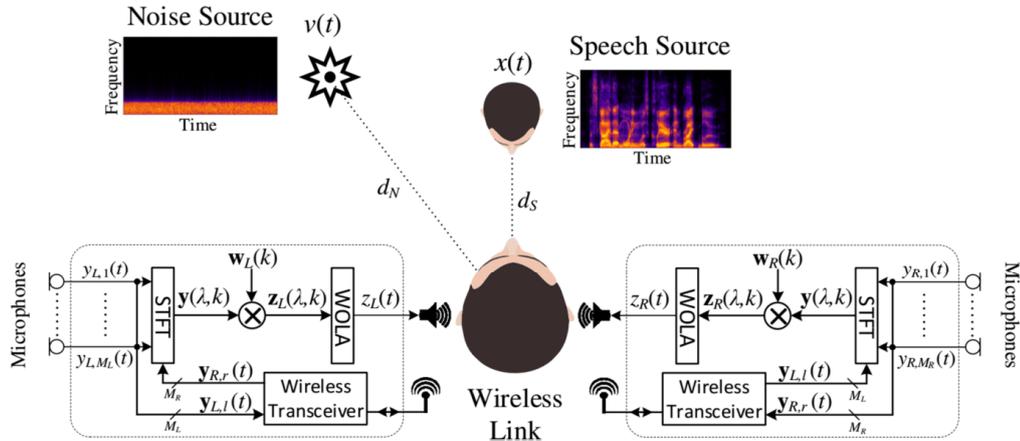


Fig. 1. Binaural system setup.

As shown in Fig. 1 the output signals of the hearing aids are

$$\begin{aligned} z_L(\lambda, k) &= \mathbf{w}_L^H(\lambda, k) \mathbf{y}(\lambda, k) \\ z_R(\lambda, k) &= \mathbf{w}_R^H(\lambda, k) \mathbf{y}(\lambda, k), \end{aligned} \quad (4)$$

where $\mathbf{w}_L(\lambda, k)$ and $\mathbf{w}_R(\lambda, k)$ are the left and right coefficient vectors of the noise reduction multichannel filter, both with dimension $M \times 1$.

III. MULTICHANNEL WIENER FILTER

The binaural MWF has been largely studied in the noise reduction context for hearing aid applications. Its cost function is given by [32]

$$J_W(k) = \mathbb{E} \left\{ \left\| \begin{bmatrix} x_{L,ref}(\lambda, k) - \mathbf{w}_L^H(k) \mathbf{y}(\lambda, k) \\ x_{R,ref}(\lambda, k) - \mathbf{w}_R^H(k) \mathbf{y}(\lambda, k) \end{bmatrix} \right\|^2 \right\}, \quad (5)$$

where $\mathbb{E}\{\cdot\}$ indicates the expected value and $\|\cdot\|^2$ is the squared Euclidean norm. Manipulating (5) leads to [33]

$$\begin{aligned} J_W(k) &= \mathbf{q}_L^T \Phi_{xx}(k) \mathbf{q}_L + \mathbf{q}_R^T \Phi_{xx}(k) \mathbf{q}_R - \mathbf{q}_L^T \Phi_{xx}(k) \mathbf{w}_L(k) \\ &\quad - \mathbf{q}_R^T \Phi_{xx}(k) \mathbf{w}_R(k) - \mathbf{w}_L^H(k) \Phi_{xx}(k) \mathbf{q}_L \\ &\quad - \mathbf{w}_R^H(k) \Phi_{xx}(k) \mathbf{q}_R + \mathbf{w}_L^H(k) \Phi_{yy}(k) \mathbf{w}_L(k) \\ &\quad + \mathbf{w}_R^H(k) \Phi_{yy}(k) \mathbf{w}_R(k), \end{aligned} \quad (6)$$

in which coherence matrices $\Phi_{xx}(k) = \mathbb{E}\{\mathbf{x}(l, k) \mathbf{x}^H(l, k)\}$ and $\Phi_{yy}(k) = \mathbb{E}\{\mathbf{y}(l, k) \mathbf{y}^H(l, k)\}$ are assumed Hermitian positive semi-definite. Equation (6) is a quadratic function of the coefficient vectors $\mathbf{w}_L(k)$ and $\mathbf{w}_R(k)$. Due to its strict convexity, the minimum of $J_W(k)$ is found in closed form by equating its partial derivatives to zero with respect to the coefficients. It was shown that the use of the obtained coefficient vectors in the system shown in Fig. 1 provides significant noise reduction and speech source spatial preservation [34].

IV. BINAURAL COST FUNCTIONS

It was previously discussed that the binaural MWF distorts the perception of the noise source localization [14]. In order to provide a trade-off between noise reduction and spatial preservation, auxiliary cost functions have been proposed in the literature [9]–[11], [30]. These cost functions are combined with $J_W(k)$ and can be generalized by:

$$J(k) = J_W(k) + \sum_i \alpha_i(k) J_i^v(k), \quad (7)$$

in which $J(k)$ is the cost function to be minimized with respect to $\mathbf{w}_R(k)$ and $\mathbf{w}_L(k)$; $J_W(k)$ is the multichannel Wiener filter cost function, responsible for noise reduction; $J_i^v(k)$ is a set of auxiliary cost functions which aim to preserve the noise binaural cues; and $i \in \{\text{ITD}, \text{ILD}, \text{ITF}\}$ for a directional noise source or $i \in \{\text{IC}\}$ for diffuse noise. Parameters $\alpha_i(k)$ are frequency dependent weighting factors that take into consideration the importance of preservation of binaural cues as compared to the noise reduction effort. Each auxiliary cost function is defined as the difference between input and output estimates of a given binaural cue. Optimization techniques are applied to (7) for finding the optimum (left/right) coefficient vectors that minimize $J(k)$ for each bin.

A. Interaural Time Difference

The ITD (in seconds) was defined in [35] as the phase of the ratio between the left and right signal components in the reference microphone. The input noise ITD at each bin and time-frame is defined as

$$\begin{aligned} ITD_{in}^v(\lambda, k) &= \frac{1}{2\pi f_s k} IPD_{in}^v(\lambda, k) \\ &= \frac{1}{2\pi f_s k} \left(\angle \frac{v_{L,ref}(\lambda, k)}{v_{R,ref}(\lambda, k)} + 2\pi p(k) \right), \end{aligned} \quad (8)$$

where \angle is the phase of its argument. The integer $p(k)$ is the phase unwrapping factor, which is unknown *a priori*, since the angle of the ratio of the spectra is computed modulo 2π . This makes the phase ambiguous above 1500 Hz due to the size and shape of the human head. For frequencies below 1500 Hz, $p(k)$ can be considered zero [35]. Under this condition, the mean input noise IPD (in radians) at a given time-window can be calculated by the following approximation

$$\begin{aligned} IPD_{in}^v(k) &= \mathbb{E} \left\{ \angle \frac{v_{L,ref}(\lambda, k)}{v_{R,ref}(\lambda, k)} \times \frac{v_{R,ref}^*(\lambda, k)}{v_{L,ref}^*(\lambda, k)} \right\} \\ &= \mathbb{E} \left\{ \angle v_{L,ref}(\lambda, k) v_{R,ref}^*(\lambda, k) \right\} \\ &\cong \angle \mathbb{E} \left\{ v_{L,ref}(\lambda, k) v_{R,ref}^*(\lambda, k) \right\} \\ &= \angle \mathbf{q}_L^T \mathbf{\Phi}_{vv}(k) \mathbf{q}_R, \end{aligned} \quad (9)$$

where $\mathbf{\Phi}_{vv}(k) = \mathbb{E} \{ \mathbf{v}(\lambda, k) \mathbf{v}^H(\lambda, k) \}$. Using the same approach for $z_L(\lambda, k)$ and $z_R(\lambda, k)$ results in the mean output noise IPD:

$$IPD_{out}^v(k) = \angle \mathbf{w}_L^H(k) \mathbf{\Phi}_{vv}(k) \mathbf{w}_R(k), \quad (10)$$

which is defined as the phase difference between the output signals in both speakers. Using (9) and (10), the ITD cost function is defined as:

$$\begin{aligned} J_{ITD}^v(k) &= |ITD_{out}^v(k) - ITD_{in}^v(k)|^2 \\ &= \frac{1}{4\pi^2 f_s^2 k^2} |IPD_{out}^v(k) - IPD_{in}^v(k)|^2 \\ &= \frac{1}{4\pi^2 f_s^2 k^2} J_{IPD}^v(k), \end{aligned} \quad (11)$$

resulting in the MWF-ITD cost function [10]:

$$J_T(k) = J_W(k) + \alpha_T(k) J_{IPD}^v(k), \quad (12)$$

in which the constant $(2\pi f_s k)^{-2}$ was included into $\alpha_T(k)$.

V. PROPOSED METHOD

In this section, we firstly analyze the accuracy of the mean IPD as an estimator of the spatial azimuth for a directional acoustic source. Following this, we propose the use of the IC for preserving the original spatialization of directional noise sources in MWF based noise reduction systems for hearing aids.

A. Performance Analysis of the IPD Estimator

It is reasonable to assume by the central limit theorem [36] that, in the STFT domain, the input noise components at both reference microphones, $v_{L,ref}$ and $v_{R,ref}$, for a given bin and time-frame, are zero-mean complex random variables, normally distributed, with coherence matrix given by the following:

$$\mathbf{\Phi}_{v_L v_R} = \begin{bmatrix} \sigma_{v_L}^2 & \rho \sigma_{v_L} \sigma_{v_R} \\ \rho^* \sigma_{v_L} \sigma_{v_R} & \sigma_{v_R}^2 \end{bmatrix}, \quad (13)$$

in which $\sigma_{v_L}^2 = \mathbb{E} \{|v_{L,ref}|^2\}$, $\sigma_{v_R}^2 = \mathbb{E} \{|v_{R,ref}|^2\}$, and $\rho = \mathbb{E} \{v_{L,ref} v_{R,ref}^*\} / (\sigma_{v_L} \sigma_{v_R})$ is the complex coherence coefficient of the left and right reference microphones. The probability density function of the random variable $\theta = \angle v_{L,ref} / v_{R,ref}$

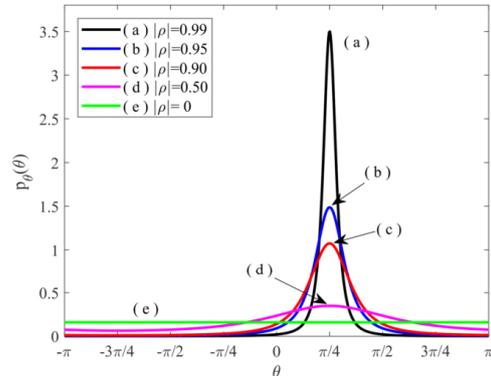


Fig. 2. Marginal probability density function for the phase of $v_{L,ref}/v_{R,ref}$ as a function of $|\rho| = |\mathbb{E}\{v_{L,ref} v_{R,ref}^*\}| / (\sigma_{v_L} \sigma_{v_R})$.

(which is equal to $IPD_{in}^v(\lambda, k)$ for $\omega < 1500$ Hz) is given by:

$$p_\theta(\theta) = \frac{1 - |\rho|^2}{2\pi(1 - \eta^2)} \left[\frac{\eta}{\sqrt{1 - \eta^2}} \arccos(-\eta) + 1 \right], \quad (14)$$

in which $\eta = |\rho| \cos(\angle \rho - \theta)$ (see Appendix).

Fig. 2 plots Eq. (14) for $\rho = |\rho| \exp(j\pi/4)$ and different values of $|\rho|$. It can be observed that the central tendency ($\pi/4$) remains fixed, but the dispersion increases with the decrease of the absolute value of the coherence coefficient $|\rho|$. In low reverberation scenarios, directional noise captured by the reference microphones naturally presents a large coherence coefficient (usually near one). Therefore, $IPD_{in}^v(k)$ results in accurate estimates for $\mathbb{E}\{\theta\}$. On the other hand, MWF-processed signals may present small coherence coefficients (as will be shown later). In this situation, $IPD_{out}^v(k)$ estimates may present large variance. In the limiting case, in which there is no coherence between residual noise components in $z_L(\lambda, k)$ and $z_R(\lambda, k)$ ($|\rho(k)| \rightarrow 0$), (14) turns to [37]:

$$p_\theta(\theta) = 1/2\pi. \quad (15)$$

In this situation, the phase has a uniform distribution. Therefore, the performance of the estimator depends on the magnitude of the coherence coefficient ρ .

B. Interaural Coherence

In consonance with results presented in Fig. 2, experiments presented in [38] indicated a strong relationship between IC (specifically to $|\rho|$) and the capacity of listeners to discriminate small ITD changes. Therefore, the authors concluded that IC should be considered cooperatively with the ITD to improve the localization of sounds in free field.

Recently, in [30], the MWF-IC technique was proposed to provide noise reduction, while preserving the dispersive characteristic of diffuse sound fields. Its cost function was defined

as follows:

$$J_C(k) = J_W(k) + \alpha_C(k) J_{IC}^v(k), \quad (16)$$

where

$$J_{IC}^v(k) = |IC_{out}^v(k) - IC_{in}^v(k)|^2, \quad (17)$$

in which the mean input and output noise IC are respectively defined as

$$IC_{in}^v(k) = \frac{\mathbf{q}_L^T \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{q}_L \mathbf{q}_R^T \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{q}_R}}, \quad (18)$$

and

$$IC_{out}^v(k) = \frac{\mathbf{w}_L^H(k) \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{w}_R(k)}{\sqrt{\mathbf{w}_L^H(k) \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{w}_L(k) \mathbf{w}_R^H(k) \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{w}_R(k)}}. \quad (19)$$

Considering a single directional noise source in free field, the noise signal can be modeled as

$$\mathbf{v}(\lambda, k) = \mathbf{h}(k) v(\lambda, k) \quad (20)$$

where $v(\lambda, k)$ denotes the noise signal, and $\mathbf{h} = [\mathbf{h}_L^T \mathbf{h}_R^T]^T = [h_{L,1}(\lambda, k) \dots h_{L,M_L}(\lambda, k) h_{R,1}(\lambda, k) \dots h_{L,M_R}(\lambda, k)]^T$ is the noise steering vector. It contains the acoustic transfer functions between the noise source and each of the M microphones. In this way, the noise coherence matrix turns to:

$$\Phi_{\mathbf{v}\mathbf{v}}(k) = \mathbb{E}\{\mathbf{v}(\lambda, k) \mathbf{v}^H(\lambda, k)\} = \sigma_v^2(k) \mathbf{h}(k) \mathbf{h}^H(k). \quad (21)$$

Using (21) in (18), after some manipulations, results in [31]

$$\begin{aligned} IC_{in}^v(k) &= \frac{\mathbf{q}_L^T \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{q}_R}{\sqrt{\mathbf{q}_L^T \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{q}_L \mathbf{q}_R^T \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{q}_R}} \\ &= \frac{\mathbf{q}_L^T \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{q}_R}{|\mathbf{q}_L^T \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{q}_R|} \\ &= e^{j \angle \mathbf{q}_L^T \Phi_{\mathbf{v}\mathbf{v}}(k) \mathbf{q}_R + 2\pi p(k)} \\ &= e^{j IPD_{in}^v(k)}, \end{aligned} \quad (22)$$

The Taylor series for the complex exponential is given by

$$e^{jx} = \sum_{k=0}^{\infty} (-1)^k \left[\frac{x^{2k}}{2k!} + j \frac{x^{2k+1}}{(2k+1)!} \right]. \quad (23)$$

Using its first order approximation in (22) it comes to:

$$IC_{in}^v(k) \cong 1 + j IPD_{in}^v(k). \quad (24)$$

Using (20) in (19) and its result and (24) in (17) leads to

$$J_{IC}^v(k) \cong |IPD_{out}^v(k) - IPD_{in}^v(k)|^2 = J_{IPD}^v(k). \quad (25)$$

Eq. (25) shows that minimization of the IC cost function for directional signals corresponds to minimization of the IPD cost function. Therefore, we proposed to apply the MWF-IC to the directional noise source case with the aim of controlling the ITD binaural cues of the processed noise.

TABLE I
OBJECTIVE MEASURES FOR THE NOISY INPUT SIGNAL

	S ₀ N ₋₃₀	S ₀ N ₋₆₀	S ₀ N ₃₀	S ₀ N ₆₀
PESQ _L	1.1	1.1	1.2	1.2
PESQ _R	1.2	1.2	1.1	1.1
SNR _L [dB]	0	0	4.6	4.5
SNR _R [dB]	4.6	4.5	0	0

VI. EXPERIMENTAL SETUP

The performance of the MWF-IC (J_C) was assessed and compared to the MWF-ITD (J_T), MWF (J_W) and unprocessed signals under objective measures and psychoacoustic experiments for one directional noise source. Simulations were performed with head-related impulse responses (HRIRs) obtained from a multichannel binaural database [39]. In this database, a manikin with the shape of a human head and torso, wearing two behind-the-ear hearing aids with three microphones on each ($M_L = M_R = 3$), was positioned inside an anechoic chamber. The acoustic sources had zero elevation, corresponding to the transverse plane of the dummy head. All acoustic scenarios presented here are comprised by one speech source and one noise source placed at distances of 80 cm and 3 m (far-field), respectively, from the manikin. The speech source was situated at zero azimuth $\theta_S = 0^\circ$, while the noise source was simulated at four different azimuths $\theta_N = \{-60^\circ, -30^\circ, 30^\circ, 60^\circ\}$ in front of the manikin. Negative azimuths correspond to the left-hand side of the sagittal plane of the manikin, while positive azimuths correspond to the opposite side.

The speech signal was a male voice selected from [40], containing a sentence of 2.7 seconds. The speech was convolved with the 0° azimuth HRIR and manually labeled to emulate an ideal voice activity detector (VAD), avoiding misclassification. The performance impact due to estimation errors of a real VAD was not approached in this study.

The noise signal was obtained by low-pass filtering white noise, to limit its energy up to 1.5 kHz (ITD range according to the duplex theory) [41], [42]. This noise signal was convolved with four different HRIRs, creating four distinct acoustic scenarios: S₀N₋₆₀, S₀N₋₃₀, S₀N₃₀ and S₀N₆₀.³ The signal to noise ratio (SNR) of the contaminated signal was defined as 0 dB in the ear closest to the noise source (called ‘‘worse ear’’), for all scenarios and experiments. The SNR and Perceptual Evaluation of Speech Quality (PESQ) of the noisy (raw) speech for the four studied acoustic scenarios are presented in Table I.

The sampling frequency was set to $f_s = 16$ kHz, and the input signals were transformed to the frequency domain by an $N = 256$ bin Short-Time Fourier Transform, with an analysis window of 128 samples, zero padding, and 50% of overlap. The transformed signals in the STFT domain were reconstructed by the weighted overlap-and-add method [43].

³S₀N_SN_{θ_N} means the speech source (S) is placed at θ_S degrees of azimuth and the noise source (N) is at θ_N degrees of azimuth with respect to the head midline (right azimuths are considered positive and left negative).

Coherence matrices $\Phi_{yy}(k)$ and $\Phi_{vv}(k)$ were computed *a-priori*, directly from the contaminated signals by applying the ideal VAD. This was done to avoid reinforcement of the binaural cues due to estimation errors [15] and to obtain an upper bound performance. Speech correlation matrices were calculated as $\Phi_{xx}(k) = \Phi_{yy}(k) - \Phi_{vv}(k)$.

The noisy speech, speech, and noise signals were processed by the coefficients obtained by applying the Broyden-Fletcher-Goldfarb-Shanno quasi-Newton optimization method [44], [45] to J_C and J_T cost functions, as well as by the theoretical solution to J_W . The weighting factor defined for a given technique is kept fixed for all bins at $\alpha(k) = \alpha$.

The experiments with volunteers were approved by the Ethics Committee in Human Research, under certificate 49741615.2.0000.0121 CEP-UFSC. All volunteers involved read and signed the written informed consent form.

A. Objective Measures

Five objective measures were calculated for evaluating the performance of both MWF-IC and MWF-ITD methods: the signal to noise ratio (SNR), which measures the noise reduction; the intelligibility weighted gain in signal to noise ratio (Δ ISNR) [46], which estimates the intelligibility of the speech signal; wideband perceptual evaluation of speech quality (PESQ) [47], which measures the overall quality of the enhanced speech signal [48]; the interaural time difference error for both speech (Δ ITD_S) and noise (Δ ITD_N) signals [11], calculated up to 1.5 kHz, which measures the preservation of ITD; and the Mean Square Coherence Error for both speech (Δ MSC_S) and noise (Δ MSC_N) [30], which measures the coherence variation between speech/noise input and output signals. Sub-indices were added to refer to speech (S), noise (N), and left (L) and right (R) ears.

Objective results obtained for scenarios S_0N_{-30} and S_0N_{-60} were very similar, respectively, to S_0N_{30} and S_0N_{60} . For this reason, only results for S_0N_{30} and S_0N_{60} are shown.

B. Psychoacoustic Experiments

Experiments with volunteers were performed to evaluate the psychoacoustic aspects of the noise signal processed by both MWF-ITD and MWF-IC. They were conducted using a head-phone (Sennheiser HD 202) connected into a laptop.

The selected group of volunteers comprised 11 males and 4 females, aging between 19 and 39 years old (mean of 29 and standard deviation of 4.9 years). No previous complaints regarding hearing losses were declared. The experimental procedure was divided into three phases: (a) learning, (b) training, and (c) testing. In the first phase, the volunteers listened to the noise (only) signal filtered by HRIRs related to seven different azimuths $\{-90^\circ, -60^\circ, -30^\circ, 0^\circ, 30^\circ, 60^\circ, 90^\circ\}$. Each audio was synchronized with visual information (presented on the laptop screen) related to the true azimuth of the processed noise signal. In the training phase, the volunteers were asked to identify the azimuth of the same seven audios, presented in random order, without previous knowledge about the true azimuths. In this phase, volunteers who performed hemisphere inversions (later-

TABLE II
WEIGHTING FACTORS FOR A MAXIMUM NOISE REDUCTION LOSS OF 15% AS COMPARED TO THE MWF TECHNIQUE

	$\alpha_T (J_T)$	$\alpha_C (J_C)$
S_0N_{-30}/S_0N_{30}	3×10^3	0.8
S_0N_{-60}/S_0N_{60}	4×10^2	0.4

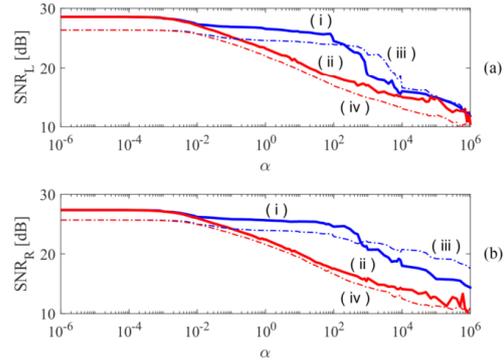


Fig. 3. SNR at the left (a) and right (b) ears for MWF-ITD (blue) and MWF-IC (red). S_0N_{60} (thick continuous line) and S_0N_{30} (thin dash-dotted line) scenarios: (i) J_T for S_0N_{60} ; (ii) J_C for S_0N_{60} ; (iii) J_T for S_0N_{30} ; (iv) J_C for S_0N_{30} .

alization errors related to the left-right sides) were dropped from the experiment. In the test phase, the remaining volunteers were requested to classify a set of 16 audios, presented in random order. A virtual protractor with 13 combo boxes, ranged from -90° to 90° in steps of 15° , was presented on the laptop screen. The selected audios comprised four noise signals presented in the training phase ($-60^\circ, -30^\circ, 30^\circ, 60^\circ$), as well as filtered versions of the noise presented in the noisy speech, for the S_0N_{-30} , S_0N_{30} , S_0N_{-60} and S_0N_{60} scenarios, according to the optimum coefficients obtained from J_W , J_T and J_C .

For performance comparison purposes, a maximum noise reduction loss of 15%, as compared to the MWF solution, was deliberately set for both MWF-IC and MWF-ITD techniques in the “worse ear”. This room establishes an arbitrary trade-off between noise reduction and spatial preservation. The weighting factors for attaining such specification, for both J_C and J_T , are presented in Table II.

VII. RESULTS

In this section, objective measures and psychoacoustic results are presented to assess the noise ITD preservation performance for both MWF-IC and MWF-ITD methods assuming the case of a single directional speech source and single directional noise source.

A. Objective Measures

Fig. 3 presents left and right SNR, as a function of the weighting factor (α), for both MWF-IC and MWF-ITD for the S_0N_{30}

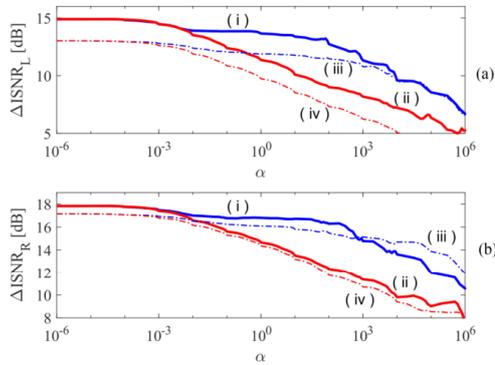


Fig. 4. Δ ISNR at the left (a) and right (b) ears for MWF-ITD (blue) and MWF-IC (red), and S_0N_{60} (thick continuous line) and S_0N_{30} (thin dash-dotted line) scenarios: (i) J_T for S_0N_{60} ; (ii) J_C for S_0N_{60} ; (iii) J_T for S_0N_{30} ; (iv) J_C for S_0N_{30} .

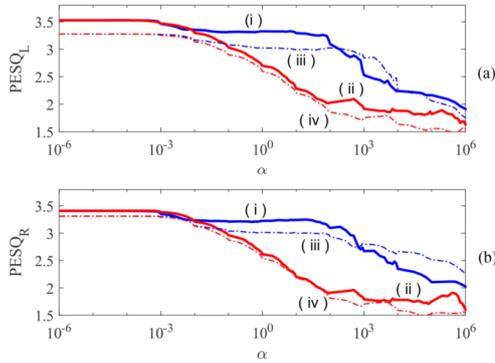


Fig. 5. PESQ at the left (a) and right (b) ears for MWF-ITD (blue) and MWF-IC (red), and S_0N_{60} (thick continuous line) and S_0N_{30} (thin dash-dotted line) scenarios: (i) J_T for S_0N_{60} ; (ii) J_C for S_0N_{60} ; (iii) J_T for S_0N_{30} ; (iv) J_C for S_0N_{30} .

and S_0N_{60} scenarios. Clearly, for both techniques, the SNR decreases with the increase of the weighting factor. The MWF-IC curves began to decrease for a smaller α as compared to the MWF-ITD, which could lead to prematurely disqualifying the former. The plateau in the extreme left side of Fig. 3 corresponds to the SNR provided by the MWF ($\alpha \rightarrow 0$) technique. The same behavior was observed for both the Δ ISNR and PESQ, respectively presented in Figs. 4 and 5.

Fig. 6a presents Δ ITD_N, which indicates that, as compared to the conventional MWF, both MWF-IC and MWF-ITD are capable of decreasing the input-output variation of the interaural time difference for increased weighting factors.

Fig. 6b shows that increasing the weighting factor of the MWF-ITD did not consistently reduce the mean square coherence error. On the other hand, MWF-IC significantly reduced

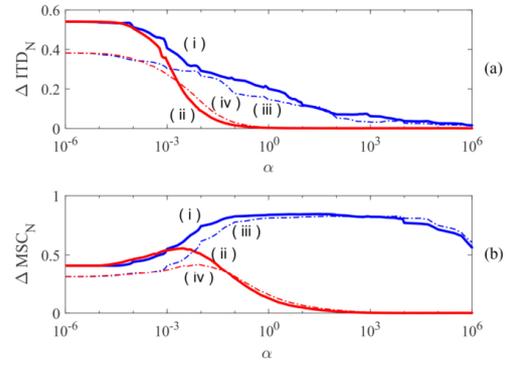


Fig. 6. Noise Δ ITD (a) and noise Δ MSC (b) for MWF-ITD (blue) and MWF-IC (red), and S_0N_{60} (thick continuous line) and S_0N_{30} (thin dash-dotted line) scenarios: (i) J_T for S_0N_{60} ; (ii) J_C for S_0N_{60} ; (iii) J_T for S_0N_{30} ; (iv) J_C for S_0N_{30} .

TABLE III
OBJECTIVE MEASURES FOR S_0N_{30} : $\alpha_T = 3 \times 10^3$ (J_T), $\alpha_C = 0.8$ (J_C)

	J_w	J_T	J_C	
SNR _L [dB]	26.3	21.3	22	
SNR _R [dB]	24.7	21.6	20.9	
Δ ISNR _L [dB]	13.4	10.6	10.3	
Δ ISNR _R [dB]	16.6	15.0	14.3	
S_0N_{30}	PESQ _L	3.4	2.8	2.7
	PESQ _R	2.9	2.8	2.4
	Δ ITD _S	2×10^{-2}	5×10^{-2}	2×10^{-2}
	Δ ITD _N	0.4	3×10^{-2}	4×10^{-3}
	Δ MSC _S	7×10^{-3}	5×10^{-3}	2×10^{-2}
	Δ MSC _N	0.4	0.82	0.1

it, restoring the original interaural coherence of the residual noise.

B. Psychoacoustic Experiments

Tables III and IV present the objective measures calculated for MWF, MWF-ITD and MWF-IC of the assessed scenarios (S_0N_{30} and S_0N_{60}). In Fig. 7, the spectrograms of the analyzed signals, for the “better ear” and S_0N_{60} scenario, are presented.

Fig. 8 shows results for the training phase of the 15 volunteers in the form of box-and-whisker diagrams. None of the volunteers performed lateralization inversions during the training phase. The box is represented by the blue rectangle that limits the upper and lower quartiles denominated q_1 and q_3 (25% and 75% of values) respectively, while the red line inside the box means the median. In addition, the box represents the interquartile range (IQR), a measure of statistical dispersion between q_3 and q_1 ($IQR = q_3 - q_1$). The whiskers show the lowest and highest sample values represented by the black dashed line. The outliers are reproduced with red crosses

TABLE IV
OBJECTIVE MEASURES FOR S_0N_{60} : $\alpha_T = 400 (J_T)$, $\alpha_C = 0.4 (J_C)$

	J_W	J_T	J_C	
SNR_l [dB]	28.5	22.7	24.2	
SNR_r [dB]	26.3	23.5	22.9	
$\Delta ISNR_l$ [dB]	14.7	12.1	12.0	
$\Delta ISNR_r$ [dB]	17.8	15.7	15.6	
S_0N_{60}	PESQ _L	3.6	2.9	3
	PESQ _R	3.1	3	2.7
	ΔITD_s	2×10^{-2}	5×10^{-2}	3×10^{-2}
	ΔITD_N	0.43	7×10^{-2}	1×10^{-2}
	ΔMSC_s	1×10^{-2}	1×10^{-2}	2×10^{-2}
ΔMSC_N	0.43	0.82	0.2	

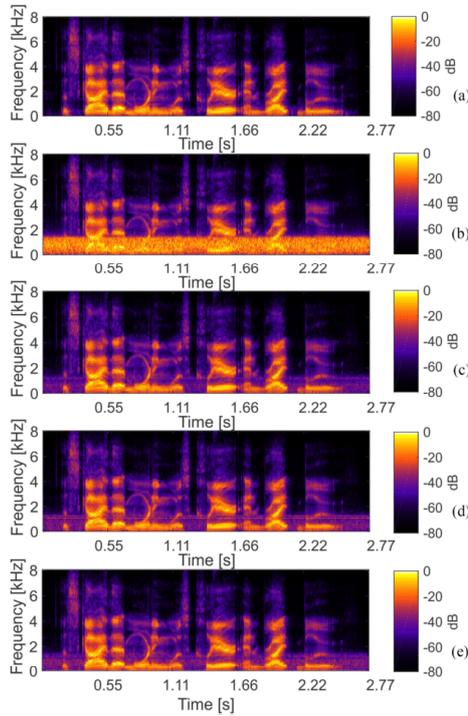


Fig. 7. Spectrograms of signals in the right ear for the S_0N_{60} scenario: (a) clean speech; (b) contaminated speech; (c) MWF processed; (d) MWF-ITD processed; (e) MWF-IC processed.

corresponding to the samples greater than $q_3 + \varpi(q_3 - q_1)$ and lower than $q_3 - \varpi(q_3 - q_1)$. The variable ϖ is set as the default value of 1.5 [49] and represents the upper and lower extremes, which are not considered outliers.

Figs. 9 and 10 show the results with the 15 volunteers in the testing phase for the S_0N_{-30} , S_0N_{30} , S_0N_{-60} and S_0N_{60}

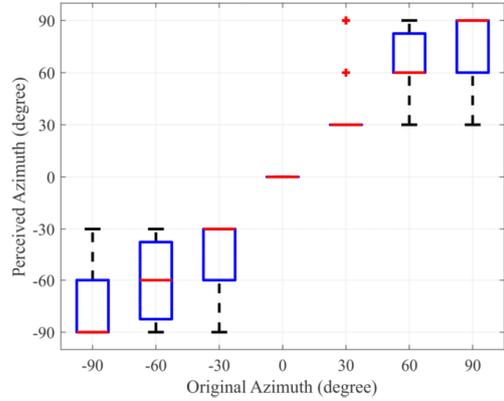


Fig. 8. Classification results obtained in the training phase for all 15 volunteers and low-pass noise.

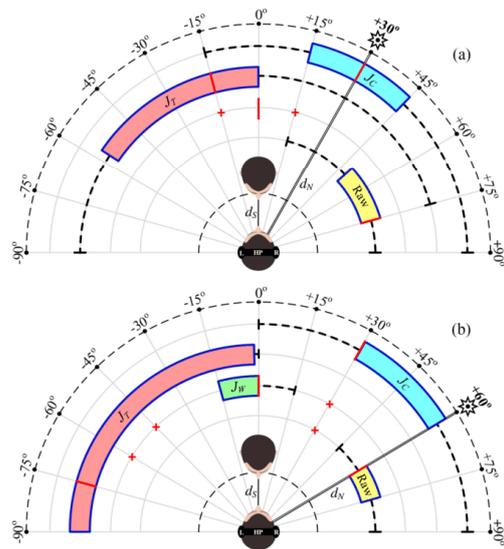


Fig. 9. Modified boxplot for the psychoacoustic experiment. Azimuth perception for the noise source. (a) S_0N_{30} (top), and (b) S_0N_{60} (bottom): Processed noise due to the MWF-IC (blue), MWF-ITD (red), MWF (green), and unprocessed noise (yellow).

scenarios. Modified box-and-whisker diagrams are presented for ease of interpretation. This innovative design embodies the spatial perception into the box-and-whisker framework. Here, the volunteer location is represented by a sketch of a human head (top view) wearing a headphone. The speech source is represented by another sketch of human head in front of the volunteer ($\theta_s = 0^\circ$) at d_s meters. The noise source is indicated by a

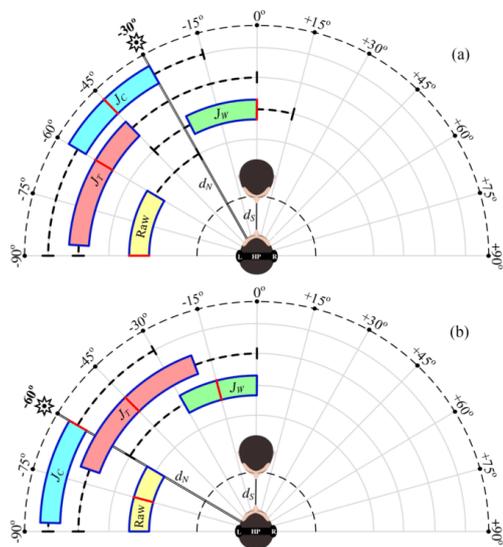


Fig. 10. Modified boxplot for the psychoacoustic experiment. Azimuth perception for the noise source. (a) S_0N_{-30} (top), and (b) S_0N_{-60} (bottom): Processed noise due to the MWF-IC (blue), MWF-ITD (red), MWF (green), and unprocessed noise (yellow).

TABLE V
PSYCHOACOUSTIC EXPERIMENT: AZIMUTH SAMPLE MEAN (\bar{x}_i), SAMPLE STANDARD DEVIATION (s_i) AND SAMPLE MEDIAN (\tilde{x}_i) FOR $i = \{Raw, W, T, C\}$, RESPECTIVELY MEANING: UNPROCESSED (RAW) AND PROCESSED BY MWF, MWF-ITD, AND MWF-IC TECHNIQUES

	S_0N_{-30}	S_0N_{30}	S_0N_{-60}	S_0N_{60}
RAW	\bar{x}_R	-74°	65°	-75°
	s_R	21°	21°	12.7°
	\tilde{x}_R	-90°	75°	-75°
J_W	\bar{x}_W	-11°	1°	-18°
	s_W	19.2°	9°	19°
	\tilde{x}_W	0°	0°	-15°
J_T	\bar{x}_T	-59°	-27°	-44°
	s_T	28.6°	44°	30°
	\tilde{x}_T	-60°	-15°	-45°
J_C	\bar{x}_C	-48°	33°	-67°
	s_C	22.1°	27.1°	19.5°
	\tilde{x}_C	-45°	30°	-60°

specific symbol with the corresponding true azimuth information (bold). The box-and-whisker diagrams are represented by boxes filled with different colors according to the legends.

Table V shows the sample mean, sample standard deviation and sample median for boxplots presented in Figs. 9 and 10. Table VI shows the absolute differences between mean and median for the MWF-ITD and MWF-IC as compared to the original noise, for all scenarios.

TABLE VI
PSYCHOACOUSTIC EXPERIMENT: ABSOLUTE DIFFERENCES BETWEEN MEAN (\bar{x}_i) AND MEDIAN (\tilde{x}_i) AZIMUTHS FOR THE MWF-ITD (T) AND MWF-IC (C) AS COMPARED TO THE PERCEIVED AZIMUTH OF THE RAW (R) NOISE

	S_0N_{-30}	S_0N_{30}	S_0N_{-60}	S_0N_{60}	Average
J_R	$ \bar{x}_R - \tilde{x}_T $	15°	92°	31°	115°
	$ \tilde{x}_R - \tilde{x}_T $	30°	90°	30°	135°
J_C	$ \bar{x}_R - \tilde{x}_C $	26°	32°	8°	21°
	$ \tilde{x}_R - \tilde{x}_C $	45°	45°	15°	30°

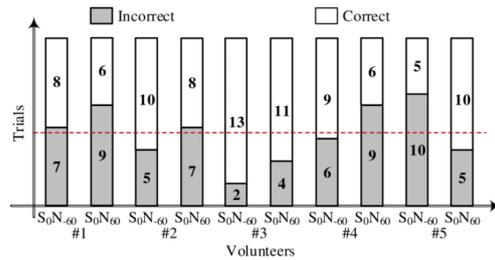


Fig. 11. Lateralization inversions obtained in the psychoacoustic experiments for the MWF-ITD technique: five volunteers and 15 trials.

The most impressive results were obtained with the MWF-ITD technique in both S_0N_{30} and S_0N_{60} scenarios. According to Fig. 9, nearly all volunteers performed lateralization inversions and, consequently, the sample means/medians pointed to the opposite hemisphere of the true localization of the noise source. Aiming to clarify these results a new experiment was performed with the five best evaluators classified in the training phase. Firstly, they listened to 15 different realizations of the noise processed by the MWF-ITD and MWF-IC techniques for the S_0N_{60} and S_0N_{-60} scenarios. Following this, they classified each signal according to the perceived hemisphere (L-left and R-right). Again, the volunteers did not perform any lateralization inversions for the MWF-IC processed signals. Furthermore, all volunteers presented hemisphere inversions for the MWF-ITD processed signals. Fig. 11 shows the results obtained for the MWF-ITD technique, in which the red dashed line indicates the average number of lateralization errors (6.4) for the 15 trials.

VIII. DISCUSSION

The objective measures SNR, $\Delta ISNR$, PESQ, and ΔITD_N presented in Figs. 3–6a clearly show a trade-off between noise reduction and spatial preservation as a function of the weighting factor. It means that acoustic comfort, intelligibility and quality are degraded, while the original noise ITD cues are preserved, with the increase of α , for both MWF-ITD and MWF-IC methods. However, the ΔMSC_N , shown in Fig. 6b, indicates that the MWF and MWF-ITD techniques may substantially change the interaural coherence of processed signals, independently of the choice of α . Conversely, the MWF-IC consistently decreased both ΔITD_N and ΔMSC_N when α increased.

Tables III and IV present objective measures for both MWF-ITD and MWF-IC, considering weighting factors designed for a maximum noise reduction loss of 15% in the “worse ear” as compared to the MWF. The SNR_R was decreased by 13% (MWF-ITD) and 15% (MWF-IC) for the S_0N_{30} scenario ($\alpha_T = 3 \times 10^3$, $\alpha_C = 0.8$), and 11% (MWF-ITD) and 13% (MWF-IC) for S_0N_{60} ($\alpha_T = 400$, $\alpha_C = 0.4$). The difference between the resulting SNRs for both MWF-ITD and MWF-IC methods did not exceed the just-noticeable difference of 3 dB [50]. In the same way, both MWF-ITD and MWF-IC presented approximately the same quality in the “worse ear” for both scenarios, since differences smaller than 0.2 PESQ are not clearly noticeable by volunteers [51]. The speech binaural cues were kept undistorted (ΔITD_S and $\Delta\text{MSC}_S < 0.1$), while ΔITD_N was considerably reduced from 0.4 to less than 0.1. The most significant difference between MWF-ITD and MWF-IC was related to the ΔMSC_N . The ΔMSC_N of the MWF processed noise, which was originally calculated as 0.4 was increased to 0.8 by the MWF-ITD, while the MWF-IC decreased it to 0.1 for the S_0N_{30} scenario and to 0.2 for the S_0N_{60} . This indicates that differently from the MWF-ITD, the MWF-IC method controlled both ITD and IC binaural cues.

The spectrogram in Fig. 7 provides a view of the “better ear”. It shows that considerable noise reduction is achieved by MWF, MWF-ITD and MWF-IC, without significant visual differences.

Fig. 8 attests that all 15 volunteers presented adequate lateralization judgment capacity. The calculated sample median accurately agreed with the true azimuths, and no hemisphere inversions were performed.

The main results of the psychoacoustic experiments are shown in Figs. 9 and 10. As expected, the MWF results (green boxes) were biased to the speech source azimuth (0°) [14]. It is also possible to verify that the median azimuth of the unprocessed noise (yellow box) is biased to a bigger azimuth magnitude as compared to its true value. This may be explained by the natural agreement between ILD and ITD binaural cues, which may amplify the correct lateralization perception. This phenomenon does not occur with the processed noise, since the analyzed techniques do not control ILD.

Two main observations can be emphasized: firstly, the MWF-IC provided more accurate (median) and precise (IQR) estimates of the true azimuths as compared to MWF-ITD (see Table V), for all acoustic scenarios. Secondly, while the MWF-IC always provided correct hemisphere localization, the MWF-ITD resulted in a significant number of hemisphere inversions, which is corroborated by Fig. 11.

The importance of IC for effective ITD cue preservation was previously analyzed in [38]. This work assessed the lateralization capacity of volunteers in reverberant environments for different values of $|IC_{in}^v|$. Acoustic noise in the effective band of the ITD (< 1.5 kHz) was investigated. Approximately 50% of hemisphere inversions were observed for scenarios with $|IC_{in}^v| < 0.2$, whereas no inversions were reported for $|IC_{in}^v| > 0.8$. These results show the relationship between the magnitude of the interaural coherence and the human ability to lateralize sounds in free-field using ITD information. An equivalent phenomenon can be observed in the MWF-IC results

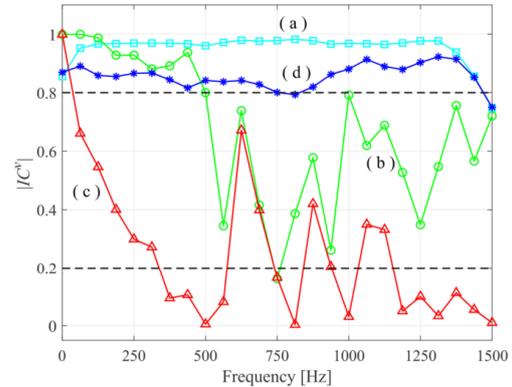


Fig. 12. Magnitude of the interaural coherence: (a) unprocessed noise (cyan square); (b) MWF processed noise (green circle); (c) MWF-ITD processed noise (red triangle); (d) MWF-IC processed noise (blue asterisk).

presented here. Fig. 12 shows an example of the magnitude of the interaural coherence as a function of the frequency for the noise signal applied in our experiments. The large coherence magnitude of unprocessed noise $|IC_{in}^v(k)|$ (cyan line - \square) is characteristic of a free-field scenario (anechoic environment). The MWF-IC processed noise (blue line - $*$) preserved the original magnitude of the interaural coherence, basically resulting in $|IC_{out}^v(k)| > 0.8$. However, both MWF and MWF-ITD techniques resulted in reduced magnitude coherences. In fact, the MWF-ITD results indicated very small values, achieving $|IC_{out}^v(k)| < 0.2$ for high frequencies. This range of IC magnitudes is characteristic of diffuse acoustic fields, like those found in highly reverberant environments. From the theoretical point of view, the coherence magnitude of the output noise should be unitary, due to the linear characteristic of the filtering process. The observed decrease in $|IC_{out}^v|$ for the MWF and MWF-ITD processed noise is probably due to amplification of STFT estimation errors by the optimum coefficient vectors $\mathbf{w}_L(k)$ and $\mathbf{w}_R(k)$. In this way, $\Phi_{vv}(k) = \mathbf{h}_v(k)\mathbf{h}_v^H(k)\sigma_v^2(k) + \Delta$, in which $\mathbf{h}_v(k)$ are the HRIRs related to the noise source localization and Δ is a full rank (Hermitian) error matrix [15]. The dashed dark lines show the limit values for the magnitude of the interaural coherence described in [38].

In [52], psychoacoustic lateralization experiments indicated that small IC magnitudes increased the importance of ILD even at frequencies below 1.5 kHz. Frequencies containing dissonant ILD and ITD information may contribute to lateralization errors. The large number of hemisphere inversions verified with the MWF-ITD, may be attributed to the combination of small magnitude interaural coherence and discordant ILD and ITD binaural cues.

Extensive additional experiments were performed, resulting in similar behavior as compared to the provided examples. In reverberant conditions ($\text{RT}_{60} = 300$ ms), psychoacoustic experiments corroborated the spatial preservation capacity of the

MWF-IC observed in anechoic conditions (outdoor environment). The main difference from the anechoic experiments is that the objective measures SNR, Δ ISNR, and PESQ are almost insensitive to changes in α due to the expected overall noise reduction degradation.

The obtained results indicate that MWF-IC is a promising technique for noise reduction with spatial preservation for both diffuse [53] and directional noise sources.

Finally, the simultaneous use of IC and ILD cost functions in (7) may lead to a noise reduction technique with perceptually relevant preservation of the spatial scenario over the entire hearing frequency range. This will be investigated in future work.

IX. CONCLUSION

This paper proposes the use of interaural coherence as a method for obtaining perceptually relevant preservation of interaural time differences in binaural hearing aids. It was shown that the MWF-ITD technique distorted the natural coherence presented by directional noise acquired by the microphones in free-field environments. The MWF-IC method provides a trade-off between noise reduction and preservation of the original noise spatialization for frequencies below 1.5 kHz. This occurs due to the simultaneous preservation of both ITD and IC binaural cues. Objective measures and psychoacoustic experiments corroborate the theoretical analysis, indicating that IC preservation is fundamental for ITD subjective perception and correct lateralization of directional sound sources. These results are of special interest to hearing aid designers in search of a binaural noise-reduction technique that preserves the original acoustic scenario for both diffuse and directional noise sources at frequencies below 1.5 kHz.

APPENDIX

Considering the ratio of two correlated circular-symmetric complex normal random variables

$$\psi = \psi_r + j\psi_i = x/y, \quad (26)$$

with zero mean and complex correlation coefficient $\rho = \rho_r + j\rho_i = \mathbb{E}\{xy\}/(\sigma_x\sigma_y)$, $\sigma_x^2 = \mathbb{E}\{|x|^2\}$, $\sigma_y^2 = \mathbb{E}\{|y|^2\}$, their joint probability density function (PDF) in rectangular coordinates is defined as [37]

$$p_{\psi_r, \psi_i}(\psi_r, \psi_i) = \frac{1 - |\rho|^2}{\pi\sigma_x^2\sigma_y^2} \left(\frac{|\psi|^2}{\sigma_x^2} + \frac{1}{\sigma_y^2} - 2\frac{\rho_r\psi_r + \rho_i\psi_i}{\sigma_x\sigma_y} \right)^{-2} \quad (27)$$

Using polar coordinates

$$\psi_r = |\psi|\cos\theta, \quad \psi_i = |\psi|\sin\theta, \quad (28)$$

in which θ denotes the phase of ψ . The joint PDF in polar coordinates is given by

$$p_{|\psi|, \theta}(|\psi|, \theta) = |\psi| p_{\psi_r, \psi_i}(\psi_r, \psi_i). \quad (29)$$

The marginal PDF of the phase is obtained by integrating (29) with respect to $|\psi|$

$$p_\theta(\theta) = \int_0^\infty p_{|\psi|, \theta} d|\psi|, \quad (30)$$

resulting in

$$p_\theta(\theta) = \frac{1 - |\rho|^2}{2\pi(1 - \eta^2)} \left[\frac{\eta}{\sqrt{1 - \eta^2}} \arccos(-\eta) + 1 \right], \quad (31)$$

in which

$$\eta = |\rho| \cos(\angle\rho - \theta). \quad (32)$$

REFERENCES

- [1] N. A. Whitmal, J. C. Rutledge, and J. Cohen, "Reducing correlated noise in digital hearing aids," *IEEE Eng. Med. Biol. Mag.*, vol. 15, no. 5, pp. 88–96, Sep/Oct. 1996.
- [2] V. Hamacher, "Comparison of advanced monaural and binaural noise reduction algorithms for hearing aids," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2002, vol. IV, pp. 4008–4011.
- [3] S. Kochkin, "MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing," *Hear. J.*, vol. 63, no. 1, pp. 19–32, 2010.
- [4] T. Bogaert *et al.*, "Horizontal localization with bilateral hearing aids: Without is better than with," *J. Acoust. Soc. Amer.*, vol. 119, no. 1, pp. 515–526, 2006.
- [5] J. Benesty, J. Chen, and Y. Huang, "Binaural noise reduction in the time domain with a stereo setup," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 8, pp. 2260–2272, Nov. 2011.
- [6] J. Li *et al.*, "Two-stage binaural speech enhancement with Wiener filter for high-quality speech communication," *Speech Commun.*, vol. 53, pp. 677–689, 2011.
- [7] H. Saruwatari *et al.*, "Hearing aid using sound-localization-preserved MMSE STSA estimator with ICA-based noise estimation," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2010, pp. 1–4.
- [8] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 18–30, Mar. 2015.
- [9] T. Bogaert *et al.*, "The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids," *J. Acoust. Soc. Amer.*, vol. 124, no. 1, pp. 484–497, 2008.
- [10] S. Doclo *et al.*, "Extension of the multi-channel Wiener filter with localization cues for noise reduction in binaural hearing aids," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2005, pp. 221–224.
- [11] T. Van den Bogaert, J. Wouters, S. Doclo, and M. Moonen, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2007, vol. IV, pp. 565–568.
- [12] T. J. Klases, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural multichannel Wiener filtering for hearing-aids: Preserving interaural time and level-differences," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006, vol. V, pp. 145–148.
- [13] S. Eyndhoven, T. Francart, and A. Bertrand, "EEG-informed attended speaker extraction from recorded speech mixtures with application in neuro-steered hearing prostheses," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 5, pp. 1045–1056, May 2017.
- [14] B. Cornelis, S. Doclo, T. Van dan Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 342–355, Feb. 2010.
- [15] B. Cornelis, M. Moonen, and J. Wouters, "Performance analysis of multi-channel Wiener filter-based noise reduction in hearing aids under second order statistics estimation errors," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 5, pp. 1368–1381, Jul. 2011.
- [16] S. Doclo and M. Moonen, "On the output SNR of the speech-distortion weighted multichannel Wiener filter," *IEEE Signal Process. Lett.*, vol. 12, no. 12, pp. 809–811, Dec. 2005.
- [17] M. L. Hawley, R. Y. Litovsky, and J. F. Culling, "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," *J. Acoust. Soc. Amer.*, vol. 15, no. 2, pp. 833–843, 2004.
- [18] T. J. Klases, T. Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.

- [19] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multi-microphone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.
- [20] V. Benichoux, M. Rébillat, and R. Brette, "On the variations of inter-aural time differences (ITDs) with frequency," in *Proc. Acoust.*, 2012, pp. 1–9.
- [21] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Amer.*, vol. 91, no. 3, pp. 1648–1661, 1992.
- [22] T. Francart, A. Lenssen, and J. Wouters, "Enhancement of interaural level differences improves sound localization in bimodal hearing," *J. Acoust. Soc. Amer.*, vol. 130, no. 5, pp. 2817–2826, 2011.
- [23] V. Willert, J. Eggert, J. Adamy, R. Stahl, and E. Korner, "A probabilistic model for binaural sound localization," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 36, no. 5, pp. 982–994, Oct. 2006.
- [24] J. Middlebrooks and D. Green, "Sound localization by human listeners," *Annu. Rev. Psychol.*, vol. 42, no. 1, pp. 135–159, 1991.
- [25] W. Gaik, "Combined evaluation of interaural time and intensity differences: Psychoacoustic results and computer modeling," *J. Acoust. Soc. Amer.*, vol. 94, no. 1, pp. 98–110, 1993.
- [26] M. H. Costa and P. A. Naylor, "ILD preservation in the multichannel Wiener filter for binaural hearing aid applications," in *Proc. Eur. Signal Process. Conf.*, 2014, pp. 636–640.
- [27] K. Blum, G.-J. Rooyen, and H. A. Engelbrecht, "Spatial audio to assist speaker identification in telephony," in *Proc. Int. Conf. Syst. Signal Image Process.*, 2010, pp. 1–4.
- [28] S. Doclo, R. Dong, T. J. Klaseen, J. Wouters, S. Haykin, and M. Moonen, "Extension of the multi-channel Wiener filter with ITD cues for noise reduction in binaural hearing aids," in *Proc. Workshop Appl. Signal Process. Audio Acoust.*, 2005, pp. 70–73.
- [29] W. M. Hartmann and E. J. Macaulay, "Anatomical limits in interaural time differences: An ecological perspective," *Frontiers Neurosci.*, vol. 8, no. 34, pp. 1–13, 2014.
- [30] D. Marquardt, V. Hohmann, and S. Doclo, "Coherence preservation in multi-channel Wiener filtering based noise reduction for binaural hearing aids," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2013, pp. 8648–8652.
- [31] E. Hadad, S. Doclo, and S. Gannot, "The binaural LCMV beamformer and its performance analysis," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 3, pp. 543–558, Mar. 2016.
- [32] T. J. Klaseen, M. Moonen, T. Van den Bogaert, and J. Wouters, "Preservation of interaural time delay for binaural hearing aids through multi-channel Wiener filtering based noise reduction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2005, pp. 29–32.
- [33] D. Wang and G. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Hoboken, NJ, USA: Wiley-IEEE Press, 2006, p. 155.
- [34] S. Doclo *et al.*, "Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions," in *Proc. Int. Workshop Acoust. Echo Noise Control*, 2006, pp. 1–4.
- [35] M. Raspaud, H. Viste, and G. Evangelista, "Binaural source localization by joint estimation of ILD and ITD," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 1, pp. 68–77, Jan. 2010.
- [36] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, 4th ed. New York, NY, USA: McGraw-Hill, 2002.
- [37] W.-J. Yan and W.-X. Ren, "Circularly-symmetric complex normal ratio distribution for scalar transmissibility functions. Part I: Fundamentals," *Mech. Syst. Signal Process.*, vol. 80, pp. 58–77, 2016.
- [38] B. Rakerd and W. M. Hartmann, "Localization of sound in rooms. V. Binaural coherence and human sensitivity to interaural time differences in noise," *J. Acoust. Soc. Amer.*, vol. 128, no. 5, pp. 3052–3063, 2010.
- [39] H. Kayser *et al.*, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP J. Adv. Signal Process.*, vol. 2009, 2009, Art. no. 298605.
- [40] Recommendation ITU-T P.50, Telephone Transmission Quality, Telephone Installations, Local Line Networks: Objective Measuring Apparatus—Artificial Voices, Appendix I: Test Signals, 1998.
- [41] T. T. Sandel *et al.*, "Localization of sound from single and paired sources," *J. Acoust. Soc. Amer.*, vol. 27, no. 5, pp. 842–852, 1955.
- [42] A. W. Mills, "Lateralization of high-frequency tones," *J. Acoust. Soc. Amer.*, vol. 32, no. 1, pp. 132–134, 1960.
- [43] R. E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 1, pp. 99–102, Feb. 1980.
- [44] J. S. Arora, *Introduction to Optimum Design*, 2nd ed. New York, NY, USA: Elsevier, 2004.
- [45] E. Habets and P. A. Naylor, "An online quasi-Newton algorithm for blind SIMO identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2010, pp. 2662–2665.
- [46] Y. Chisaki *et al.*, "Real-time processing using the frequency domain binaural model," *Appl. Acoust.*, vol. 68, pp. 923–938, 2007.
- [47] J. V. Michalowicz *et al.*, "A general Isserlis theorem for mixed-Gaussian random variables," *Statist. Probab. Lett.*, vol. 81, pp. 1233–1240, 2011.
- [48] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.
- [49] D. Shannon, "Box-and-whisker plots with the SAS," *Pharmaceutical Statist.*, vol. 2, pp. 291–295, 2003.
- [50] D. McShefferty, W. M. Whitmer, and M. A. Akeroyd, "The just-meaningful difference in speech-to-noise ratio," *Trends Hear.*, vol. 20, pp. 1–11, 2006.
- [51] A. Servetti and J. C. Martin, "Error tolerant MAC extension for speech communications over 802.11 WLANs," in *Proc. Veh. Technol. Conf.*, 2005, pp. 1–5.
- [52] W. M. Hartmann, B. Rakerd, and Z. D. Crawford, "Transaural experiments and a revised duplex theory for the localization of low-frequency tones," *J. Acoust. Soc. Amer.*, vol. 139, no. 2, pp. 968–985, 2016.
- [53] D. Marquardt and S. Doclo, "Interaural coherence preservation for binaural noise reduction using partial noise estimation and spectral postfiltering," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 7, pp. 1261–1274, Jul. 2018.



and low-power DSP architectures.



and vibration control.

Fábio Pires Itturriet received the B.E.E. degree from Universidade Católica de Pelotas, Pelotas, Brazil, in 2008, and the M.Sc. degree in microelectronics from Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, in 2012. He is currently working toward the Ph.D. degree with Universidade Federal de Santa Catarina, Florianópolis, Brazil. He is currently a Professor of Automation and Control Engineering with the Instituto Federal Sul-Rio-Grandense, Charqueadas, Brazil. His research interest is in digital signal processing, including binaural hearing aids

Márcio Holsbach Costa received the B.E.E. degree from Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, in 1991, the M.Sc. degree in biomedical engineering from Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, in 1994, and the Dr. degree in electrical engineering from Universidade Federal de Santa Catarina, Florianópolis, Brazil, in 2001. From 1994 to 2004, he was with the Department of Electrical Engineering, Biomedical Engineering Group, Universidade Católica de Pelotas, Pelotas, Brazil. Since 2004, he has been with the Department of Electrical and Electronic Engineering, Universidade Federal de Santa Catarina. In 2013, he was a Visiting Researcher with the Communications and Signal Processing Research Group, Imperial College London. His present research interests are in biomedical signal processing, hearing aids, linear and nonlinear adaptive filters, adaptive inverse control and active noise and